Universidad Autónoma de Madrid
Programa de Doctorado en Biociencias Moleculares

Papel de los Telomeric repeat-containing RNA (TERRA) en la biología del telómero y en pluripotencia mediado por la proteína telomérica TRF1

Tesis Doctoral

Juan José Montero Valderrama

Madrid, 2019

Universidad Autónoma de Madrid
Facultad de Ciencias
Departamento de Biología Molecular

FACULTAD DE CIENCIAS

Papel de los Telomeric repeat-containing RNA (TERRA) en la biología del telómero y en pluripotencia mediado por la proteína telomérica TRF1

Tesis Doctoral

Juan José Montero Valderrama
BSc, MSc

Todo el trabajo presentado en esta tesis ha sido llevado a cabo en el laboratorio de Telómeros y Telomerasa en el Centro Nacional de Investigaciones Oncológicas (CNIO, Madrid) bajo la dirección y supervisión de la Dr. Maria Blasco Marhuenda y la Dr. Isabel López de Silanes Asenjo

Madrid, 2019

Madrid

Adaptado mínimamente de Agustín Lara

(CHOTIS)

Cuando llegues a Madrid, chulapo mío
voy a hacerte emperador de Lavapiés;
y alfombrarte con claveles la Gran Vía,
y a bañarte con vinillo de Jerez.
En Chicote, un agasajo postinero
con la crema de la intelectualidad
y la gracia de un piropo retrechero
más castizo que la calle de Alcalá.

Madrid, Madrid, Madrid,
pedazo de la tierra en que nací
por algo te hizo Dios
la cuna del requiebro y del chotis.
Madrid, Madrid, Madrid,
en Múnich se piensa mucho en tí
por el sabor que tienen tus verbenas
por tantas cosas buenas
que soñamos desde aquí;
y vas a ver lo que es canela fina
y armar la tremolina
cuando llegues a Madrid.

A Madrid por hacerme tan feliz.

Agradecimientos

Agradecimientos

Gracias a Maria Blasco por darme la oportunidad de realizar la tesis doctoral en el laboratorio, el apoyo económico y la supervisión. Gracias a Isabel por su supervisión, las correcciones y todas las discusiones científicas.

Gracias a Nani por darme la oportunidad de trabajar con ella y los filipinos blancos. Gracias también al resto de telómeros por todos estos años, especialmente a aquellos que han realizado la tesis conmigo, Iole, MA y Leire, por su apoyo y todas las cervezas. También gracias a los ex-supresores por todos los buenos ratos pasados juntos y a Manuel Serrano por las discusiones científicas.

Gracias a todas las unidades del CNIO y a toda la gente del CNIO en general, habéis hecho de mi tiempo aquí algo inolvidable. Especialmente gracias a la unidad de confocal y a Diego por haber sido un pilar fundamental para mí y para tantos otros. También gracias a Raúl y Sandra de citogenética por todas las discusiones.

Gracias a mis compañeros de piso, actuales y pasados, Leila, Lola y Mara por haber sido como mi familia y aguantarme.

Gracias a mi gente de Madrid, Inés, Lucas, Ángel, Cristina, Sergio y a la última incorporación a mi marida Amaia. Por vuestra culpa probablemente vaya a vivir 5 años menos y necesite un trasplante de hígado, pero a vuestro lado todo esto ha merecido la pena.

Por último, gracias a mi familia, a mis padres y a mi hermana, porque sin vosotros no habría llegado hasta aquí ni hasta a ningún sitio.

Agradecimientos

Resumen/Summary

Resumen

Debido al estado altamente heterocromatinizado del telómero se pensaba que el final de los cromosomas no era transcrito. Sin embargo, hace una década se demostró que los telómeros se transcriben en una familia de lncRNAs denominados *Telomeric repeat-containing RNA* o TERRA. En los últimos años, un gran número de funciones diferentes han sido asociadas a estos transcritos. Sin embargo, la falta de modelos genéticos KO para TERRA ha impedido confirmar estas funciones en sistemas celulares. La falta de modelos se debe mayoritariamente a que su origen subtelomérico no está claro. En humanos, los estudios más recientes han propuesto que TERRA se podría transcribir de 18 *loci* diferentes, dificultando en gran medida la generación de células KO para TERRA.

Uno de los principales objetivos de mi tesis, ha sido identificar los posibles *loci* de TERRA humanos, con el objetivo de generar células humanas KO para TERRA. Primero evaluamos los 18 *loci* previamente descritos, encontrando que el 80% de estos *loci* estaban estructuralmente conservados. Sin embargo, los transcritos de estos subtelómeros estructuralmente conservados no mostraban características típicas de TERRA. Sin embargo, sí encontramos características propias de TERRA en los transcritos del subtelómero 20q. Esto nos permitió generar células humanas tumorales ALT 20q-TERRA KO. Estas células presentaban una drástica disminución de los TERRA totales y nos permitieron probar que TERRA es esencial para mantener la integridad y la longitud telomérica. En una siguiente aproximación, usando de nuevo el modelo de células 20q-TERRA KO, caracterizamos el papel de TERRA en la formación de la heterocromatina telomérica. Así, demostramos que TERRA es esencial para el depósito de marcas de heterocromática, incluyendo a H3K27me3 (una marca que no se había descrito previamente en el telómero) y HP1 a los telómeros. El papel de TERRA en el mantenimiento de la heterocromatina telomérica estaba mediado por su interacción con el complejo PRC2 y el reclutamiento del mismo a los telómeros.

Por otra parta, hemos caracterizado el mecanismo por el cual la proteína telomérica TRF1 es importante para el mantenimiento de la pluripotencia y el papel de TERRA en este proceso. TRF1 es esencial para la protección telomérica, previniendo que los telómeros se fusionen y degraden, además diferentes evidencias experimentales apuntan a que TRF1 tiene también un papel importante en el mantenimiento de la pluripotencia. Encontramos que el silenciamiento de TRF1 en iPSCs (células madre pluripotentes inducidas) murinas causa un gran cambio en la expresión génica, fundamentalmente de genes asociados a pluripotencia y diferenciación celular. Estos cambios estaban asociados a un dramático incremento de la unión de PRC2 al DNA y de la deposición de la marca del sintetizada por este complejo, la H3K27me3. Hemos demostrado que el silenciamiento de TRF1 incrementa los niveles de TERRA y que TERRA se une a los promotores de los genes cuya expresión cambia tras eliminar TRF1. Estos resultados son consistentes con un modelo en el que los cambios de expresión de TERRA dependiente de TRF1 modulan el reclutamiento de PRC2 (*Polycomb Repressive Complex 2*) y la expresión de genes asociados a pluripotencia y diferenciación.

Summary

Due to the highly heterochromatic status of the telomere, it was thought that the telomere not transcribed. However, a decade ago, our lab found telomere transcription. Telomere transcription gives rise to a family of lncRNAs called Telomeric repeat-containing RNA or TERRA. Over the last few years, a plethora of different functions have been assigned to TERRA, but the *in vivo* role of TERRA remain unknown. This is due to the lack of TERRA KO genetic as a result of the lack of information about the TERRA *locus*. In human, a study has proposed that TERRA could be transcribed from 18 different *loci*. That large amounts of loci still avoiding the generation of cells lacking TERRA. However, in mice, it has been described that TERRA are generated mostly from some subtelomeres.

Based on this, the main aim of my thesis was to identify the human TERRA *loci* in order to generate human cells lacking TERRA. To this end, we evaluated the 18 different *loci* previously proposed to be TERRA origins. We found that 80% of the *loci* were in structurally conserved subtelomeres. We could not find TERRA features for the transcripts arising from these structurally conserved subtelomeres. Interestingly, we found TERRA features in the transcripts from the subtelomere 20q. Next, generated human ALT cancer cells knock-out for the 20q-TERRA *locus* by using the CRISPR-Cas9 system. These cells presented a drastic downregulation of total TERRA. The 20q-TERRA KO cells allowed us to characterize the TERRA role in telomere biology and we found that TERRA is essential to maintain telomere integrity and telomere length. Then, we characterized the role of TERRA in telomere heterochromatin formation. We proved that TERRA is essential for the deposition of heterochromatic histone marks including H3K27me3, a mark that is not normally associated with the telomere, and HP1 to the telomeres. We found that the role of TERRA in telomeric heterochromatin maintenance could be mediated through their interaction with the Polycomb repressive complex 2 (PRC2) and its recruitment to the telomeres.

Next, we characterized the mechanism of the telomeric protein TRF1 in pluripotency maintenance and the implication of TERRA in this phenomenon. TRF1 is essential for telomere protection, preventing telomeres from degradation and fusion. There are evidences suggesting that TRF1 could have a role in the maintenance of pluripotency and embryonic development. We found that TRF1 depletion in murine iPSCs (induced pluripotent stem cells) cause drastic changes in gene expression, mostly in genes related with pluripotency maintenance and cell differentiation. Also we found that these trascriptional changes were accompanied by vast epigenetics changes. The changes found include an increase in PRC2 binding to the DNA, and H3K27me3 deposition. We found that TRF1 depletion increases TERRA transcription and that TERRA binds to genes which expression changes after TRF1 depletion. These results are consistent with a model in which TRF1-dependent changes in TERRA levels modulate polycomb recruitment to pluripotency and differentiation genes.

Índice

Índice

Artículos

Discusión

Conclusiones

Bibliografía

Anexos

Índice

Abreviaciones

Abreviaciones

ALT	Alternative Lengthening of Telomeres
APBs	ALT-associated Promyelocytic Leukemia nuclear bodies
ATRX	Alpha-Thalassemia/Mental Retardation Syndrome, Nondeletion Type, X-Linked
CCND1	Cyclin D1
DSB	Double Strand Break
DNA	Deoxyribonucleic Acid
DNMTs	DNA Methyltransferases
DKC1	Dyskerin 1
DotL	Disruptor of telomeric Silencing-Like
ESCs	Embryonic Stem Cells
ERK	Extra-cellular signal Regulated Kinase
EXO1	*Exonuclease 1*
GSK3	Glycogen Synthase Kinase 3
H2A	Histone H2A
H2B	Histone H2B
H3	Histone H3
H3k18ac	Acetyl-Histone H3 (Lys18)
H3K27me3	Tri-Methyl-Histone H3 (Lys27)
H3K36me3	Tri-Methyl-Histone H3 (Lys36)
H3K4me3	Tri-Methyl-Histone H3 (Lys4)
H3K56ac	Acetyl-Histone H3 (Lys56)
H3K79me3	Tri-Methyl-Histone H3 (Lys79)
H3K9ac	Acetyl-Histone H3 (Lys9)
H3K9me3	Tri-Methyl-Histone H3 (Lys9)
H4	Histone H4
H4K20me3	Tri-Methyl-Histone H4 (Lys20)
HATs	Histone Acetyltransferases
HDACs	Histone Deacetylases
HMTases	Histone Methyltransferases
hnRNPs	Heterogeneous Nuclear Ribonucleoproteins
HP1	Heterochromatin Protein 1
IKK	Inhibitor of KappaB Kinase

Abreviaciones

ITSs	Interstitial Telomeric Sequences
Kb	Kilobase
KDMs	Histone Lysine Demethylases
KLF4	Kruppel-like Factor 4
KMTs	Histone Lysine Methyltransferases
KO	Knock Out
lncRNAs	Long non-coding RNAs
Lys	Lisina
m7G	7-Methylguanosine
MEFs	Mouse Embryonic Fibroblasts
MEK	MAP kinase kinase
MYC	Myelocytomatosis viral oncogene homolog
NF-κB	Nuclear Factor kB
NHP2	NHP2 ribonucleoprotein
NMD	Nonsense-mediated Decay
NOP10	*NOP10 ribonucleoprotein*
OCT4	Octamer 4
PML	Promyelocytic Leukaemia Protein
POT1	Protection of Telomeres Protein 1
PRC2	Polycomb Repressive *Complex* 2
PRDMs	Histone Arginine Demethylases
PRMTs	Histone Arginine Methyltransferases
RAP1	Repressor Activator Protein 1
Rb1	RB Transcriptional Corepressor 1
Rbl1	RB Transcriptional Corepressor Like 1
Rbl2	RB Transcriptional Corepressor Like 2
R-loop	RNA Loop
RNA	Ribonucleic Acid
RNAPII	RNA Polymerase II
SIRT6	Sirtuin 6
Suv39h	Suppressor of Variegation 3-9
Suv4-20h	Suppressor of Variegation 4-20
TERRA	Telomeric repeat–containing RNA
Terc	Telomerase RNA Component

TERT	Telomerase Reverse Transcriptase
TIN2	TRF1 Interacting Protein 2
T-Loop	Telomere Loop
TPE	Telomere Position Effect
TPP1	TINT1/PTOP/PIP1
TRF1	Telomere Repeat Binding Factor 1
TRF2	Telomere Repeat Binding Factor 2
T-SCE	Telomere Sister-Chromatid Exchange
ZSCAN4	Zinc Finger and SCAN domain containing 4
3´-OH	3'-hydroxy group

Introducción

Introducción

1. Los telómeros

1.1. Introducción, descubrimiento y estructura primaria

El termino telómero deriva de los términos griegos *telos* (final) y *meros* (parte o porción) y se refiere a una estructura nucleoproteíca altamente heterocromatinizada al final de los cromosomas eucariotas y esencial para el mantenimiento de la estabilidad cromosómica al evitar que los finales de los cromosomas sean reconocidos como una rotura de doble cadena de DNA o *double-strand breaks* (DSBs).

Los telómeros fueron descritos por primera vez por Hermman Muller y Barbara McClintock entre finales de los años 30 y principios de los 40, al estudiar los efectos de los rayos X en los cromosomas de *Drosophila Melanogaster* y de *Zea mays*. Basándose en estos estudios ellos concluyeron que el final de los cromosomas eran estructuras discretas y absolutamente necesarias para el mantenimiento de la estabilidad cromosómica [1,2]. No fue hasta 1978 cuando la secuencia de los telómeros fue descrita por Elisabeth Blackburn y Joe Gall. Ellos descubrieron que la secuencia del final de los cromosomas del protozoo *Tetrahymena thermophila* consistía en un número variable de repeticiones en tándem de 6 nucleótidos TTGGGG y que potencialmente los telómeros de todas las especies estarían compuestos por repeticiones en tándem de secuencias de oligonucleótidos [3]. Desde entonces, las secuencia teloméricas de muchos organismos han sido caracterizadas y con el tiempo ha quedado claro que en la mayoría de los casos los telómeros son estructuras altamente conservadas, constituidas por repeticiones en tándem de secuencias cortas ricas en nucleótidos de guanina [4].

Una excepción importante y probablemente la más estudiada es el caso de los telómeros de *Drosophila Melanogaster* y otras especies del género, cuyos telómeros están constituidos por copias de tres retrotransposones distribuidos de forma aleatoria a lo largo de estas estructuras [5].

1.2. El DNA telomérico

Como se mencionó anteriormente una parte muy importante de los telómeros es su secuencia constituida por repeticiones de DNA en tándem de secuencias ricas en guanina. En el caso de todos los vertebrados donde la secuencia se ha identificado, estas repeticiones consisten en la secuencia TTAGGG [6]. Al final de la doble cadena de DNA conformada por estas secuencias repetitivas se encuentra un fragmento protuberante de DNA de cadena simple que sobresale del extremo 3´. Este fragmento también denominado *G-strand overhang* (Figura 1), está compuesto por entre 150 y 200 nucleótidos y es esencial para la correcta función protectora del telómero [7,8].

El número de repeticiones de DNA en tándem en el telómero, también denominada longitud telomérica varía a lo largo de diferentes organismos y dentro de la misma especie dependiendo del estado de desarrollo, reduciéndose esta además con la edad. Es más, también es heterogénea dentro de un mismo organismo al ser dependiente del tipo celular [9,10]. La longitud telomérica media en humanos varía entre las 5 y 15 Kb y en ratón entre 15 y 40 Kb [11–13]. Esta

variación en la longitud telomérica no afecta al correcto funcionamiento del telómero, hasta que su tamaño alcance una longitud críticamente corta. Sin embargo, cuando esto ocurre los telómeros son reconocidos como DSBs y pierden su funcionalidad, culminando en un incremento de la inestabilidad genómica, la posible pérdida de información genética y en muchos casos la muerte celular por apoptosis o senescencia [14,15].

1.3. El complejo de proteínas teloméricas, shelterinas o telosoma

A pesar de la gran importancia que tiene la estructura del DNA telomérico y la longitud de la misma, esta no es suficiente para el mantenimiento telomérico, sino que requiere de la interacción con un complejo de seis proteínas denominadas shelterinas o telosoma [16,17] para la

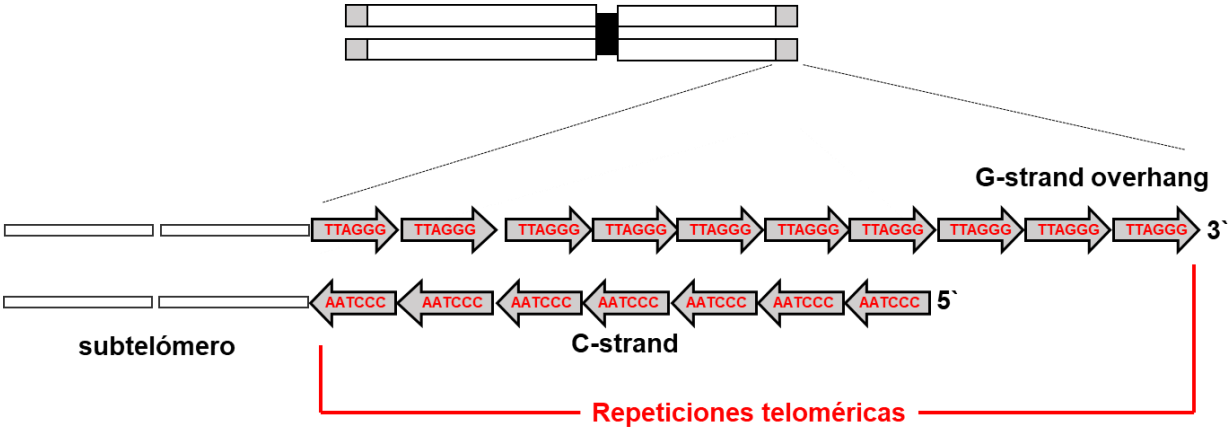


Figura 1. Estructuras de los telómeros. Representación esquemática del DNA telomérico, compuesto por repeticiones en tándem de la secuencia TTAGGG

correcta protección telomérica. Los componentes de este complejo son los siguientes: *telomere repeats factors* 1 y 2 (TRF1 y TRF2), *TRF1-interacting factor* 2 (TIN2), *protection of telomeres 1* (POT1), *POT1-TIN2 organizing protein 1* (TPP1) y *repressor/activator protein 1* (RAP1) [16,18,19].

1.3.1. Estructura del complejo shelterina

TRF1, TRF2 y POT1 son los únicos componentes de este complejo que interaccionan directamente con el DNA telomérico, TRF1 y TRF2 con cadena doble de DNA y POT1 con cadena simple (*G-strand overhang).* TRF1 y TRF2 presentan un alto grado de homología y se unen al telómero en forma de homodímeros [20,21]. Además de unirse al *G-strand overhang* POT1 está conectado al complejo de shelterinas a través de su asociación directa con TPP1 [22]. La proteína encargada de anclar a todo el complejo es TIN2, que hace interaccionar al complejo formado por TPP1/POT1 con TRF1 y TRF2 a través de la interacción TRF1, TRF2 y TPP1 [23–25]. Finalmente, RAP1 es reclutado al telómero por su asociación con TRF2. [23–25] (Figura 2). Además de estás seis proteínas que forman el núcleo del complejo shelterina, muchas otras proteínas pueden interaccionar de forma transitoria con el telómero, la mayoría de ellas a través de la

interacción con las shelterinas, sobre todo con TRF1 y TRF2 [23] y juegan un importante papel a la hora de la función protectora del telómero y su regulación.

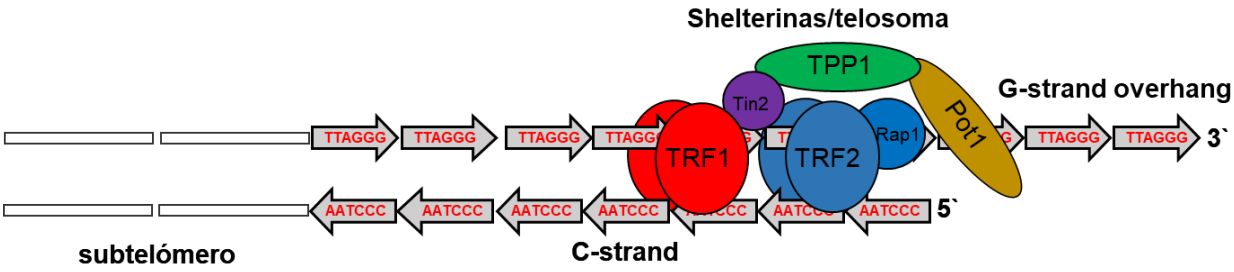


Figura 2. Estructura del complejo shelterina o telosoma.

1.3. 2. Función telomérica de las shelterinas

La correcta interacción de las shelterinas con el telómeros constituye el llamado *capping* telomérico, que es esencial para la protección del telómero, evitando su degradación y la fusión con otros cromosomas. Además, este *capping* telomérico y la acción específica de algunas de las proteínas que lo componen impide que el telómero sea reconocido como DSBs y por tanto la activación de la respuesta a daño en el DNA o *DNA damage response* (DDR) [16]. Otro papel clave de algunas shelterinas es la regulación de la longitud telomérica como pueden ser el caso de TRF1 y TRF2 [26], POT1 [27], TIN2 y TPP1 [28]. Además de esto las shelterinas pueden tener otras funciones como TRF1 que juega un rol exclusivo facilitando la replicación del telómero [29,30] o POT1 que al ser la única shelterina que interacciona con el *G-strand overhang* es esencial para el mantenimiento del mismo [31] (Tabla 1).

1.3.3. Función extratelomérica de las shelterinas

Más allá de los roles de las shelterinas derivados de su interacción directa con el telómero, algunas de ellas presentan unión a otras partes del DNA y funciones no relacionadas con el mantenimiento del telómero. Esto es en parte debido a la existencia de repeticiones teloméricas en el interior de los cromosomas, formando las llamadas secuencias teloméricas intersticiales o *Interstitial Telomeric Sequences* (ITSs) [32]. Estás regiones están especialmente enriquecidas en los subtelómeros (las regiones de DNA adyacentes al inicio de las repeticiones en tándem teloméricas), pero también están repartidas por el resto del genoma. Incluso pueden llegar a constituir grandes estructuras repetitivas normalmente asociadas a regiones heterocromatina pericentromérica en algunas especies [33,34].

Un caso destacado es el de RAP1, que a pesar de que su unión al telómero no es directa sino derivada de su interacción con TRF2 [23,24], se ha reportado que puede unirse de forma directa a regiones no teloméricas mayoritariamente a ITSs, aunque también puede unirse a regiones sin la secuencia telomérica. Estas uniones están enriquecidas en zonas subteloméricas mostrando un papel en el silenciamiento de genes subteloméricos y en el control de la expresión de genes no subteloméricos [35]. Otra función más allá de su interacción con el telómero es la de ser un

importarte modulador del factor nuclear κB (NF-κB) a través de su presencia en el citoplasma de células humanas y su interacción con el inhibidor de NF-κB IKK [36].

La presencia de TRF1 y TRF2 en ITSs también ha sido descrita en células humanas [37,38]. Además, se ha sugerido que la interacción de TRF2 con los ITSs es importante para la organización nuclear de los telómeros a través de su interacción con la proteína de la envoltura nuclear lamin A/C [39]. Sin embargo, hasta ahora la unión de TRF1 en ITTs en células de ratón no se ha podido confirmar, es más en fibroblastos embrionarios de ratón o *Mouse embryonic fibroblas* (MEFs) se ha reportado que esta no ocurre [40]. Recientemente, se ha descrito que TRF2 es capaz de unirse a otras regiones de heterocromatina difíciles de replicar de forma no canónica y sin necesidad de la secuencia telomérica. Esta unión se da especialmente en los pericentrómeros y es requerida para la replicación de la heterocromatina pericentromérica y de otras regiones heterocromáticas a lo largo de todo el genoma (Tabla 1).

2. Acortamiento telomérico y mecanismos de re-elongación telomérica

Shelterina implicada	Función telomérica
TRF1	Replicación telomérica
POT1	Protección del G-*strand overhang*
TRF1, TRF2, PPO1, RAP1	Inhibición de la respuesta por daño en el DNA
TRF1, TRF2, POT1, TIN2, TPP1	Regulación de la longitud telomérica
TRF1, TRF2, POT1, TIN2, TPP1, RAP1	*Capping* telomérico

Shelterina implicada	Función extratelomérica
TRF1, TRF2, RAP1	Unión a *Interstitial Telomeric Sequences*
RAP1	Control de la expresión génica
TRF2	Replicación y estabilidad extratelomérica
TRF2	Anclaje de los telómeros a la envoltura nuclear

Tabla 1. Función telomérica y extratelomérica de los componentes del complejo shelterina

2.1. El problema de replicar el final de los cromosomas

El correcto mantenimiento de la longitud telomérica o su correcto capping son esenciales para la estabilidad del mismo y por lo tanto, para la integridad de los cromosomas. A pesar de esto el telómero se acorta durante cada división del ciclo celular como resultado de que la replicación del DNA por parte de las polimerasas del DNA es incompleta para moléculas de DNA lineal, como es el caso de los cromosomas de eucariotas [41,42]. El mecanismo de este acortamiento se debe a la naturaleza antiparalela de las hebras de DNA, que conforman una doble cadena de DNA, y a que el mecanismo de replicación del DNA es típicamente semiconservativo y bidireccional. Las polimerasas de DNA solo pueden añadir nuevas bases a los extremos 3´de la hebra de DNA que se está sintetizando. En la cadena líder, la maquinaria de DNA puede funcionar continuamente, junto con la progresión de la horquilla de replicación en una dirección 5´-3´. En la cadena retrasada, sin embargo, las polimerasas de DNA se mueven en la dirección opuesta a la horquilla de replicación y por tanto, la síntesis de DNA se tiene que realizar de una manera discontinua. Además, en la horquilla de replicación de DNA, una vez

abierta, unas polimerasas de RNA específicas denominadas primasas han de sintetizar pequeños oligonucleótidos de RNA complementarios a la cadena de DNA parental. Estos oligos denominados cebadores, son los que generan el extremo 3´-hidroxilo (3´-OH) que las polimerasas de DNA usan como sustrato para sintetizar la cadena retrasada. Esto resulta en unos fragmentos discontinuos de DNA denominados fragmentos de Okazaki en cuyas zonas intermedias aparecen los cebadores de RNA [43]. Dichos fragmentos de RNA son luego degradados y los huecos son rellenados con su correspondiente secuencia en bases de DNA. Sin embargo, justo al extremo de los cromosomas, cuando el cebador de RNA es degradado, la polimerasa de DNA es incapaz de terminar de replicar el DNA por la ausencia de un extremo 3´-OH. Esto da como resultado en que las cadenas de DNA de nueva síntesis son más cortas con cada división celular [44–46]. Por otra parte, como la cadena líder se sintetiza de una manera completa, obteniéndose una molécula de DNA con extremos romos. Luego, el extremo 5´ de esta molécula es digerido por dos nucleasas (EXO1 and Apollo/SNM1B) para formar el *G-strand overhand*. Esto resulta en una cadena 5´ más corta y el acortamiento telomérico[47–49]. En resumen, los telómeros se acortan debido a la replicación discontinua de la cadena retrasada y la acción de nucleasas digiriendo la cadena líder.

Este fenómeno conocido con el "problema en la replicación terminal" conlleva la pérdida de entre 50 y 200pb de los telómeros en cada división celular, resultando en un acortamiento telomérico relacionado con la edad o el número de divisiones celulares. Esto es una de las principales causas del envejecimiento tanto de células como de los organismos y conlleva la muerte celular una vez que las células alcanzan una longitud telomérica críticamente corta [46]. Para resolver este problema y poder dividirse un mayor número de veces, en situaciones tanto fisiológicas como patológicas, algunas células (como pueden ser las células madre embrionarias, las células madre adultas, algunas células somáticas como los linfocitos durante la expansión monoclonal o las células cancerosas), pueden usar dos mecanismos que les permite alargar los telómeros. Estos son la acción de una enzima llamada telomerasa [50] o en ausencia de esta un mecanismo denominado alargamiento alternativo de los telómeros o *Alternative Lengthening of Telomeres* (ALT).

2.2. Telomerasa

La telomerasa es una retrotranscriptasa compuesta por una subunidad catalítica de carácter proteico llamada TERT y un componente de RNA no codificante denominado Terc que funciona como molde para la síntesis de *novo* de secuencias de repeticiones teloméricas en el telómero [50] (Figura 3). La telomerasa fue descrita por primera vez por Carol Greider y Elizabeth Blackburn en 1985 y actúa al reconocer el grupo 3´-OH del *G-strand overhang*, funcionando como una retrotranscriptasa y añadiendo repeticiones teloméricas al mismo extremo [50,51]. Algunos componentes adicionales son necesarios para estabilizar el complejo, entre ellos se encuentra fundamentalmente la proteína DKC1 y otras proteínas de unión a RNA como NOP10 y NHP2 [52–55] (Figura 3).

El ciclo de acción de la telomerasa puede dividirse en tres pases principales: 1) Reconocimiento del grupo 3´-OH del *G-strand overhang* y unión de la enzima; 2) síntesis de la primera repetición telomérica en sentido 5´-3´; 3) Translocación de la enzima y unión de nuevo al sustrato usando un nuevo 3´-OH para empezar la siguiente ronda de síntesis [56]. La telomerasa es capaz de realizar múltiples rondas de síntesis telomérica y que tiene cierto grado de procesividad [57].

La actividad de la telomerasa compensa el acortamiento telomérico de aquellas células donde se está expresando, como las células madre embrionarias, las células germinales, las células madre adultas y también en la gran mayoría de los tumores [58–62]. En las células madres adultas, sin embargo, la actividad de la telomerasa no es capaz de mantener la longitud telomérica a través de las distintas divisiones celulares, por lo tanto sus telómeros se acortan, generando uno de los principales mecanismos moleculares causantes del envejecimiento [9,44,63].

2.3. *Alternative Lengthening of Telomeres* (ALT)

Cuando la actividad telomerasa no está presente, existen algunos mecanismos menos frecuentes capaces de alargar los telómeros conocidos colectivamente como *Alternative Lengthening of Telomeres* (ALT), y que fueron descritos por primera vez en tumores [64]. Estos mecanismos están basados en procesos de recombinación homóloga del DNA, entre telómeros o telómeros y subtelómeros (Figura 3). Sin embargo, los mecanismos moleculares detrás del ALT y la iniciación de los procesos de recombinación no están del todo claros y existen diferentes modelos que tratan de explicarlo [65,66].

Una de las principales características asociadas al ALT es la presencia de telómeros con longitudes teloméricas muy heterogéneas dentro del mismo núcleo, con la presencia de telómeros extremadamente largos y telómeros cortos [67]. Otra es la asociación de los telómeros

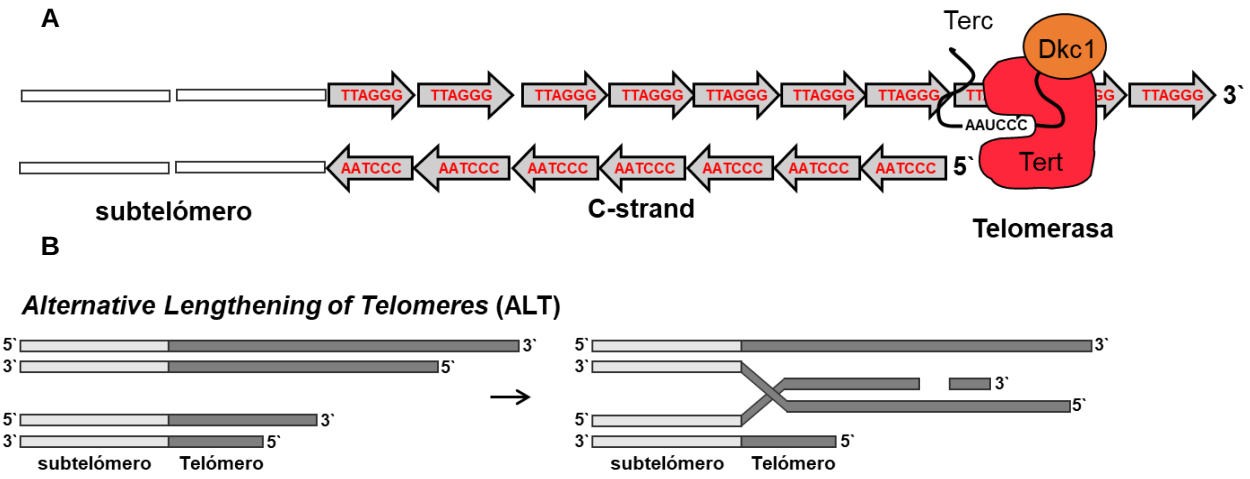


Figura 3. Mecanismos para alargar los telómeros. (A) Esquema de los componentes básicos de la telomerasa y de su unión al telómero. **(B)** Esquema del funcionamiento del Alternative Lengthening of Telomeres (ALT).

con la proteína PML, formando los llamados cuerpos PML asociados al ALT o *ALT-associated*

PML bodies (APBs) [64,67–69]. Por último, debido al proceso de recombinación durante el ALT se producen unos círculos de DNA extracromosomales o *c-circles* [67,69].

El ALT se encuentra principalmente en condiciones patológicas. Entre un 10-15% de los tumores elongan los telómeros por ALT, pero la distribución no es homogénea siendo la incidencia mayor en los sarcomas y el mecanismo de elongación telomérica mayoritario en ciertos tipos de tumores [70]. Además, muchas líneas celulares inmortalizadas muestran ALT [64,68,71]. El ALT no está restringido a situaciones patológicas, ya que también se ha observado en algunas células embrionarias en eventos tempranos de desarrollo [72] y se ha demostrado que puede ser de gran importancia en el mantenimiento de la longitud de telomérica en iPSCs y ESCs en cultivo, estando este mecanismo asociado a la expresión de Zscan4 [73,74].

3. Cromatina telomérica y estructuras secundarias del telómero

Más allá de las características descritas hasta ahora, los telómeros presenta otras capas de complejidad. Como otras regiones de DNA presentan nucleosomas [75,76] y estos están sometidos a una compleja regulación epigenética [77–81]. También debido a su particular composición los telómeros presentan diferentes estructuras secundarias [8,82]. Todos estos elementos son esenciales para el correcto mantenimiento y la funcionalidad del telómero.

3.1. Introducción a la cromatina telomérica

La cromatina es el complejo formado por el DNA y distintas proteínas. La unidad básica de la cromatina son los nucleosomas, que consiste en 146 pb de DNA envueltas alrededor de un octámero de proteínas de histona (H2A, H2B, H3 y H4). Uno de los principales mecanismos para regular la estructura de la cromatina son las modificaciones postraduccionales de las histonas y estos habitualmente se dan en los aminoácidos, lisina, arginina, serina, tirosina y treonina. Estas modificaciones *per se* pueden alterar la interacción entre las histonas y el DNA o servir como señal para el reclutamiento de complejos modificadores o remodeladores de la cromatina, resultando en cambios que pueden afectar a la compactación de la cromatina, la expresión génica u otras funciones del genoma[78,83–86].

Una de las modificaciones de las histonas más comunes e importante para el telómero es la adición de grupos metilos a una lisina [87]o a una arginina [88]. Las enzimas encargadas de realizar esta modificación son las arginina y lisina metiltransferasas (PRMTs y KMTs, respectivamente) o en conjuntos las histonas metiltransferasas (HMTases). Las encargadas de eliminarlas son las arginina y lisina desmetilasas (PRDMs y KDMs, respectivamente) [89–91]. Dependiendo del contexto cromatínico y la interacción con distintos remodeladores de histonas, esta metilación de histonas puede estar asociada con la activación o la represión de la transcripción. La metilación de histonas además juega un papel esencial en la compactación de la cromatina, formando una cromatina altamente empaquetada denominada heterocromatina. Las marcas de histonas H3K9me3, H3K27me3, h3K79me3 y H4K20me3 están asociadas a regiones altamente heterocromatinizadas [92–100].

Otra marca de histonas fundamental, también encontrada en ciertas ocasiones en los telómeros, es la adición de un grupo acetilo a una lisina. Estas marcas son establecidas por una serie de enzimas llamadas histona acetiltransferasas (HATs) y eliminadas por las denominadas histona deacetilasas (HDACs) [89–91]. La acetilación de histonas disminuye la carga positiva que se encuentra habitualmente en las mismas, debilitando la interacción de las histonas con el DNA, descompactando la cromatina y generando la denominada eucromatina [101,102].

La presencia de una estructura cromatínica altamente compactada o heterocromatinizada es una característica clásicamente asociada a los telómeros de levadura y mamíferos. Además, la heterocromatina telomérica no se encuentra limitada a las repeticiones teloméricas TTAGGG, sino que puede extenderse cientos de kilobases en dirección al centrómero, heterocromatinizando la región subtelomérica y silenciando la expresión de los genes ubicados en estas regiones. Este fenómeno es conocido como efecto de la posición telomérico o *telomere position effect* (TPE)[77,103]. Esta heterocromatina está caracterizada por la presencia de modificaciones de histona específicas y proteínas asociadas a la cromatina [77,78,80,81,103] (Figura 4).

Sin embargo, el concepto de que los telómeros son siempre estructuras altamente heterocromatinizadas está abierto a debate. Esto se debe principalmente a que estas conclusiones se basan en datos obtenidos mayoritariamente en telómeros de ratón [104], pero estudios hechos en células humanas muestran que sus telómeros están menos enriquecidos en marcas heterocromáticas de lo esperado [93,105–108]. A pesar de esto, hay evidencias experimentales directas de que la presencia de estas marcas heterocromáticas en los telómeros humanos es esencial para el correcto funcionamiento de los mismos, al igual que pasa con los murinos [109–112]. Otros datos que ponen de manifiesto que no todos los telómeros tiene que estar siempre altamente heterocromatinizados, es el hecho de que algunos organismos presentan otra serie de características. Estos pueden ser los casos de los telómeros de *Drosophila Melanogaster* y otras especies del género que, dependiendo del cromosoma, tienen telómeros con características de heterocromatina o de eucromatina [113] o el caso de los telómeros de *Arabidopsis thaliana* y otras plantas que muestran características tanto de cromatina heterocromática como eucromática [114,115]. A esto se le suma que la cromatina telomérica puede ser más dinámica de lo que se pensaba en un principio y los telómeros pueden pasar de una conformación más heterocromática a más eucromática en determinadas ocasiones como el acortamiento telomérico o la reprogramación celular para la obtención de iPSCs [10,79] (Figura 5). De estos procesos se hablará más adelante en esta sección.

3.2. Marcas y establecimiento de la heterocromatina telomérica

Las histonas de la cromatina telomérica de los mamíferos están muy poco acetiladas y muestran características similares a otros elementos génicos repetitivos y heterocromáticos como pueden ser los pericentrómeros. Además de la falta de acetilación en las histonas, la

heterocromatina telomérica está caracterizado por la presencia de las marcas H3K9me3, H3K20me3 y la presencia de la proteína de la heterocromática 1 o *Heterocromatin Protein 1* (HP1) [103,109].

Las encargadas de definir las marcas de cromatina en estas estructuras repetitivas son las HMTases Suv39h y Suv4-20h [116,117]. Los telómeros y subtelómeros están enriquecidos en la marca H3K9me3, que es sintetizada por Suv39h1 y Suv39h2. La presencia de la marca H3K9m3 genera un sitio de unión con alta afinidad para HP1, siendo uno de los primeros eventos secuenciales para la heterocromatinización del telómero [80,118,119]. Posteriormente HP1 interaccionaría con Suv4-20h1 y Suv4-20h2 y establecen la marca H4K20m3 en el telómero, completando el proceso de heterocromatinización [103,120,121] (Figura 4). Otras dos vías que favorecen el depósito de H4K20me3 en el telómero han sido descritas, siendo la primera mediada por la familia de las proteínas supresoras de tumores RB (Rb1, Rbl1 y Rbl2), que directamente interaccionarían con Suv4-20h reclutándola al telómero [122–126]. La segunda estaría mediada por la HMTase DotL, que sintetizaría la marca H3K79me2, y esta sería capaz de favorecer el establecimiento de H4K20me3 [97] (Figura 4).

Pero no solo la correcta deposición de marcas de het erocromatina es necesaria para

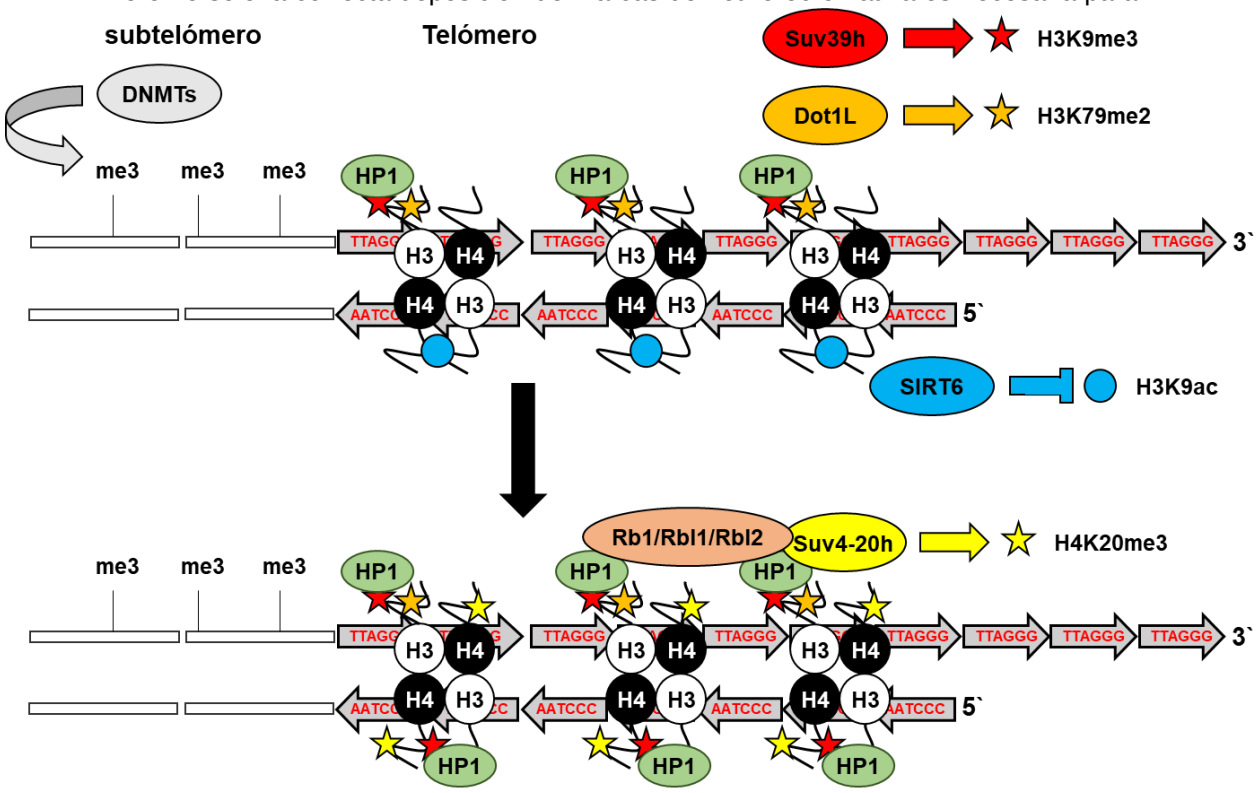


Figura 4. Proceso de heterocromatinización del telómero. Representación esquemática proceso de heterocromatinización del telómero y de las marcas epigenéticas asociadas al mismo. (Adaptada de Shoeftner and Blasco, 2010).

la correcta formación de la cromatina telomérica, como se ha mencionado antes, el telómero presenta muy poca cantidad de histonas a cetiladas y esta desacetilación también es importante para su funcionamiento. La principal responsable de esto es la HDAC SIRT6, que es capaz de

eliminar los grupos acetilos de las marcas H3K9ac, H3k18ac y H3K56ac, siendo la más frecuente en el telómero la H3K9ac [109,110,127] (Figura 4).

La última característica importante de la cromatina telomérica es la metilación de citosinas, que también ha estado clásicamente asociada con la heterocromatina y el silenciamiento de la expresión génica. Esta no puede ocurrir directamente en las repeticiones teloméricas TTAGG debido a la ausencia de parejas de CG, que es el sustrato necesario para estas metilaciones. En cambio, los subtelómeros de mamíferos están fuertemente metilados, y esta metilación es producida por las metilasas de DNA DNMT1, DNMT3a y DNMT3b [11,81,128,129] (Figura 4). A pesar de no estar en los telómeros esta metilación es importante para el mantenimiento de los mismos.

3.3. Consecuencias de la pérdida de la heterocromatina telomérica

Distintos modelos tantos de ratón *in vivo* como celulares han puesto de manifiesto la gran importancia de la correcta heterocromatinización para el correcto funcionamiento del telómero. El efecto más estudiado, es su gran importancia para la regulación de la longitud telomérica. Esto deriva de evidencias experimentales donde se han eliminado proteínas o complejos necesarios para el correcto ensamble de la heterocromatina telomérica, como puede ser el caso de las HMTases Suv39h y Suv4-20h, las proteínas de la familia Rb o también de aquellas metiltransferasas encargadas de la metilación del DNA subtelomérico [79–81,103,122,123]. Junto con la pérdida de la heterocromatina telomérica o subtelomérica derivada de eliminar estos moduladores, se observa un gran incremento de la recombinación entre cromátidas hermanas en el telómero o *Telomere Sister-Chromatid Exchange* (T-SCE), dando lugar a unos telómeros aberrantemente largos, siendo esta elongación mediada muy posiblemente por la activación del ALT [81,103]. Además, de acuerdo con la activación del ALT en estos modelos, se observa un gran incremento de los APBs, que como se comentó en la sección 2.3, es una característica importante del ALT [79–81,103,130]. A pesar de esto cabe destacar que HP1, proteína esencial para la heterocromatinización del telómero, puede jugar un doble papel en el establecimiento de ALT, al ser también esencial para la formación de los APBs [131].

Otra función importante de la cromatina telomérica es el TPE. De hecho, la eliminación de SIRT6, encargada de la desacetilación de las histonas teloméricas y subteloméricas, principalmente al estabilizar la marca H3K9ac, causa una reactivación de la expresión de los genes subteloméricos [132]. Otra evidencia es que en ESC tratadas con inhibidores de HDACs, también se observa una pérdida del TPE y una elongación de los telómeros, probablemente también asociada con el ALT como en los casos anteriores [109]. Otro factor importante para el mantenimiento del TPE es la metilación del subtelómero, ya que la eliminación de la misma también afecta negativamente a este mecanismo [76,81,133–135].

Por último, otro mecanismo potencialmente asociado a la heterocromatina telomérica sería la actividad de la telomerasa y la protección telomérica. Esto se ha abordado directamente

en células humanas, donde se ha sobreexpresado la proteína HP1, fusiona con TRF1, por lo tanto, consiguieron depositar HP1 exclusivamente en el telómero, heterocromatinizándolo de forma artificial. La consecuencia de esto es un incremento en la protección al daño en el DNA, pero también una inhibición de la actividad telomerasa en el mismo [112].

En resumen, un telómero muy heterocromatinizado o compactado, presentaría una mayor protección frente al daño y un correcto TPE, mientras que un telómero más eucromático sería más dado a la activación de ALT y más accesible para la telomerasa. Esto es especialmente interesante porque la cromatina telomérica puede cambiar entre un estado y otro dependiendo del contexto celular, como en el caso de la reprogramación celular o el acortamiento telomérico (Figura 5).

3.4. Cambios fisiológicos y patológicos en la cromatina telomérica

El acortamiento telomérico produce grandes cambios en la compactación telomérica, ESCs sin Dicer 1 presentan telómeros mucho más largos y esto se asocia con un incremento de la compactación telomérica [79]. En el caso contrario, la sobreexpresión de modulares negativos de la longitud telomérica como puede ser TRF2 en células de ratón, resulta en un acortamiento telomérico y una pérdida de las marcas asociadas a la heterocromatina telomérica [136]. Todo esto

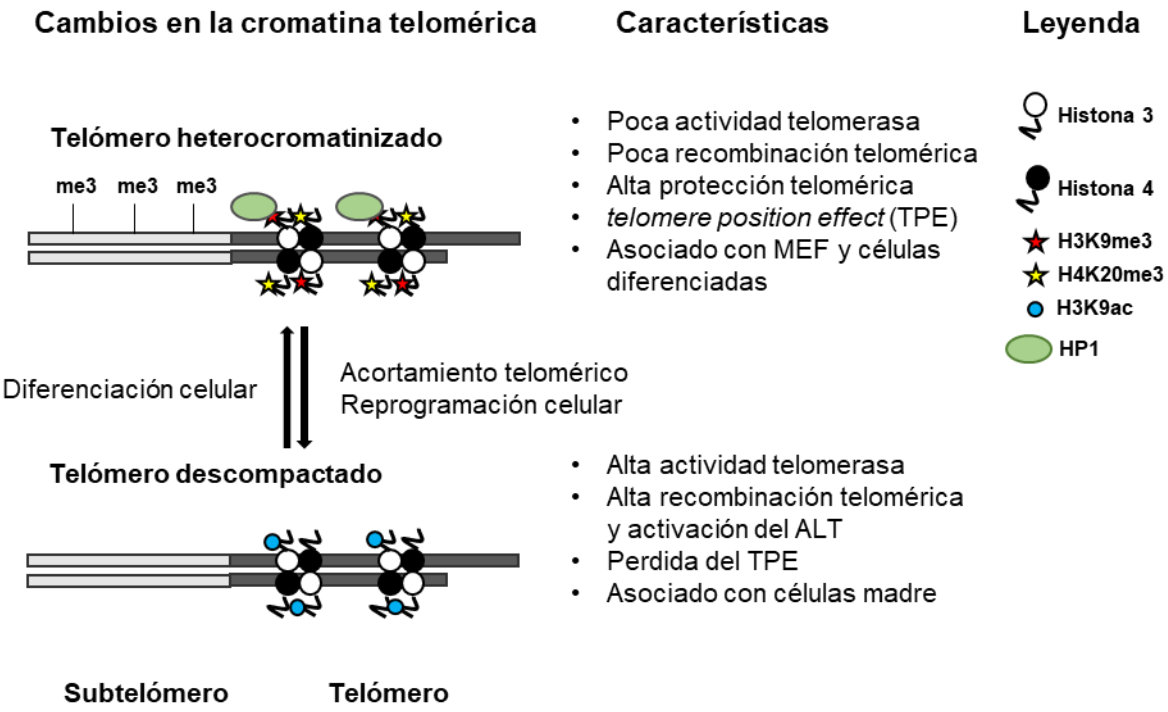


Figura 5. Cambios en la composición de la epigenética telomérica. Esquema que muestra los cambios en la epigenética del telómero tras distintos eventos celulares y sus consecuencias funcionales.

se ha abordado directamente usando células KO para TERC, estas células presentan una pérdida progresiva de la longitud telomérica, acompañada de la pérdida de las marcas de heterocromática y ganancia de H3K9ac. Todo esto a su vez correlaciona con un aumento en el número de T-SCE, asociado a un aumento de ALT [79]. Estos datos sugieren un modelo, según el cual, durante la vida de un organismo, sus telómeros se van acortando. Esto se asocia con una

pérdida de la compactación telomérica, lo que podría favorecer el rescate de esta longitud, mediante ALT y haciendo a los telómeros más accesibles para la telomerasa (Figura 5).

También se ha visto que procesos de diferenciación celular o el proceso de reprogramación (por el cual una célula puede ganar las características de una célula madre embrionaria (ESC) pluripotente generando las denominadas *induced pluripotent stem cells* (iPSCs)) pueden generar cambios en la longitud y epigenética del telómero. De hecho, la compactación telomérica es mayor en MEFs, que en las ESC de los cuales se originan. [10,103]. El proceso de reprogramación está acompañado por un proceso de elongación telomérica mediado por la telomerasa, que continua hasta que se alcanza una longitud telomérica similar a la de las ESC. Esto se acompaña con una pérdida de la cantidad de HP1, H3K9me3 y H3K20me3 en el telómero y una ganancia de H3K9ac, una composición de la cromatina telomérica más similar a la encontrada en las ESCs [10]. Esto evdencia un proceso de rejuvenecimiento de los telómeros durante el paso de MEF hasta iPSCs, en los que estos ganan una serie de características más similares a los de las ESC, incluyendo una cromatina más abierta (Figura 5).

A parte de en estas condiciones no patológicas, ciertas mutaciones en los factores necesarios para la correcta compactación de la cromatina, como la pérdida de la deposición de H3K20me3 por cambios en la expresión de Suv4-20h o de Suv39h, o la pérdida de la metilación del subtelómero se podrían encontrar en situaciones patológicas. Igual que ocurre en las células de ratón, esto podría estar asociado con una pérdida del TPE o una elongación telomérica anómala vía ALT, que podría ser importante para los eventos de tumorogénesis o tener importancia en el envejecimiento [134,137,138]. Sin embargo, debido a los grandes cambios por todo el genoma que ocurren tras la mutación en factores importantes para la remodelación de la cromatina, es difícil asociar un posible papel de los mismos meramente a su efecto sobre el telómero. A pesar de todo esto, hay ejemplos de que la pérdida de la regulación de genes controlados por TPE, puede ser importante para algunas enfermedades humanas [133,139,140].

3.5. Estructuras secundarias del telómero

Además de la cromatina telomérica, debido a sus singulares características, el telómero presenta ciertas estructuras secundarias, siendo las más importantes la formación de G-cuádruplex y el T-loop.

La cadena rica en guaninas del telómero, puedo formar unas estructuras denominadas G-cuádruplex o G4 [82]. Los G-cuádruplex son estructuras en los que cuatro guaninas se agrupan entre ellas de forma espontánea a través de unos enlaces no canónicos denominados enlaces de Hoogsteen (Figura 6). Estas estructuras solo requieren de la repetición telomérica, formándose incluso sin la presencia de shelterinas [141,142]. Aunque su funcionamiento no está claro, ha sido asociado a la protección telomérica y a la regulación de la actividad telomerasa [143].

En cambio, el llamado T-loop es una estructura de DNA y proteína que ocurre cuando el *G-strand overhand* invade la doble hebra de DNA telomérico, formando el T-loop. Para esta formación es esencial la interacción con el complejo shelterina [8,144,145] (Figura 6). Esta estructura evita que el final del telómero sea reconocido como DSB y lo protege de su degradación [8,145].

4. *Telomeric repeat-containing RNA* (TERRA)

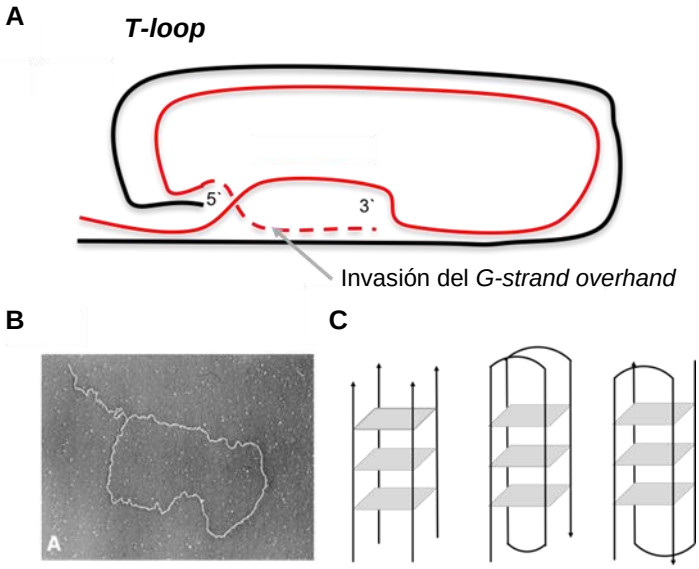


Figura 6. Estructuras secundarias del telómero. (A) Estructura del *T-loop*. **(B)** Imagen de microscopía electrónica donde se observa el *T-loop* en células HeLa (Griffith et al., 1999). **(C)** Estructural del G-cuádruplex o G4.

4.1. Introducción y características de TERRA

Debido a la alta heterocromatinización del telómero, la presencia de estructuras secundarias y el efecto que esto produce en el silenciamiento subtelomérico (TPE) [77,103], tradicionalmente se ha pensado que los telómero eran transcripcionalmente inactivos. Sin embargo, los telómeros pueden ser transcritos en unos RNAs largos no codificantes o *long non-coding RNAs* (lncRNAs), enriquecidos en la secuencia telomérica UUAGGG y por tanto denominados *Telomeric repeat-containg RNA* o TERRA. Estos transcriptos se han descrito en multitud de organismos eucariotas muy alejados evolutivamente entre ellos, como pueden ser, organismos simples como protozoos y levaduras [146–149] hasta un amplio rango de eucariotas pluricelulares; plantas, moscas [113,150] y también en vertebrados, humanos, ratones y en pez cebra [151,152], poniendo de manifiesto la gran conservación evolutiva de la transcripción telomérica. El descubrimiento de dicha transcripción telomérica supuso añadir una nueva capa de complejidad a la biología del telómero, y el estudio de la función de TERRA ha supuesto uno de los grandes retos de este campo, habiéndose asignado a TERRA multitud de funciones, no estrictamente relacionas entre ellas a lo largo de los años y difíciles de confirmar por la falta de modelos genéticos.

En mamífero, TERRA es transcrito desde la región subtelomérica hasta el telómero, es decir, en dirección centrómero-telómero, usando exclusivamente la hebra rica en G del telómero como hebra codificante y la rica en C como hebra molde [153]. Esto hace que una de las principales características de TERRA sea una secuencia de RNA híbrida, conteniendo tanto una secuencia de RNA correspondiente al subtelómero en su inicio 5´ y la repetición telomérica UUAGGG en su terminación 3´ [151,152]. Además de esto, las moléculas de TERRA son heterogéneas en tamaño, como indica su detección por Northern. Así se observan un continuo de moléculas de RNA que van desde los 100 nucleótidos hasta las 9 Kb [151,152]. (Figura 7). Estos transcritos son nucleares,

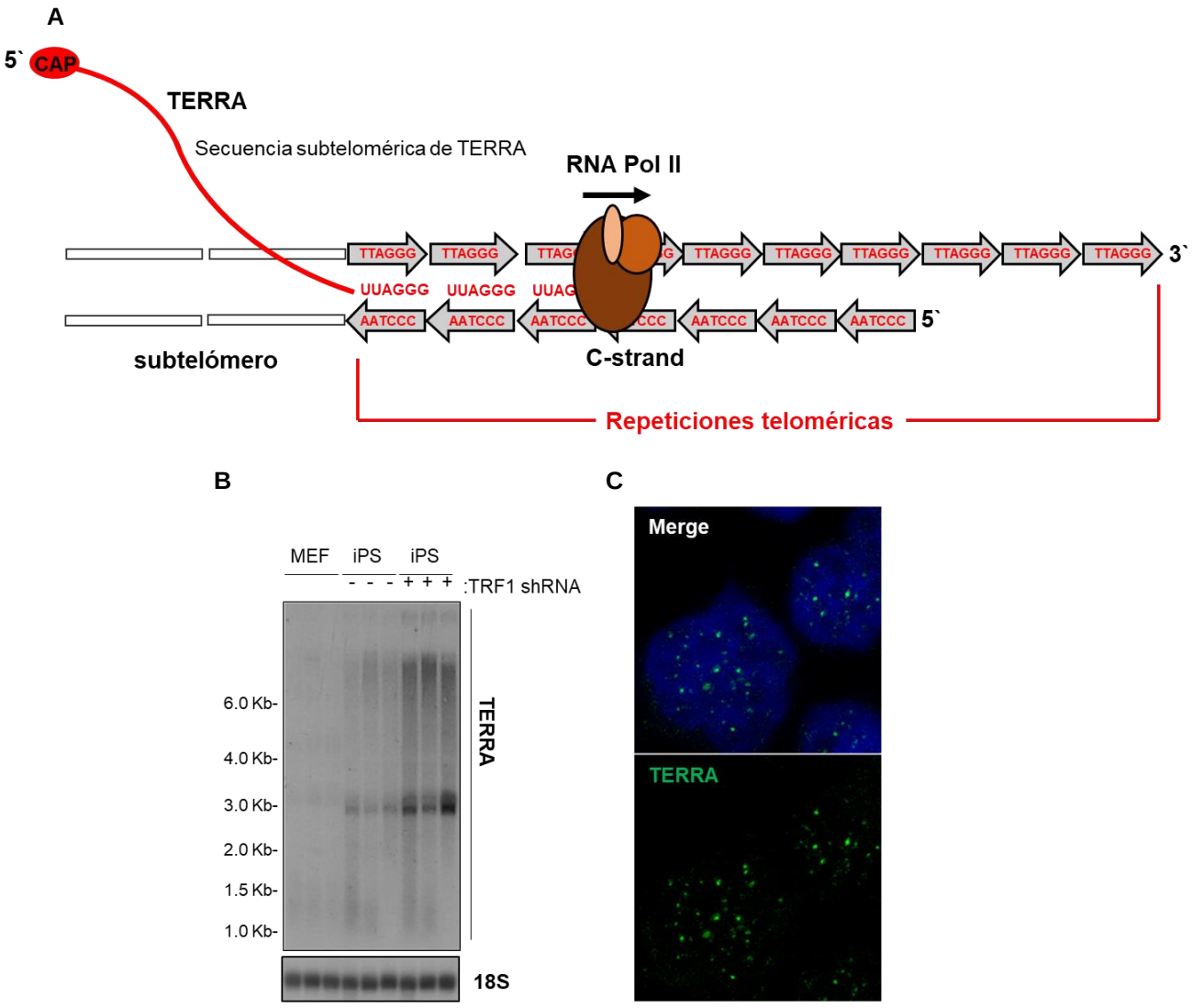


Figura 7. Telomeric repeat-containing RNA (TERRA). (A) Esquema que muestra el proceso de transcripción de TERRA a partir del telómero por la RNA polimerasa II (RNA Pol II) y sus características. **(B)** Northern Blot contra la repetición telomérico de TERRA en MEF, IPS de ratón e IPS donde se ha silenciado TRF1 (Marion, Montero et al. 2019) **(C)** Imagen de microscopía confocal de un RNA-FISH contra las repeticiones teloméricas de TERRA (en verde) en células humanas U2OS. (DAPI en azul).

y cuando se detecta por RNA-FISH revelan un patrón nuclear discreto y punteado, que co-localizan en un 30% de los casos con los telómeros [151,152] (Figura 7). Aunque otras RNA polimerasas pueden contribuir parcialmente en la expresión de TERRA, esta es llevada a cabo

de forma mayoritaria por la RNAPII, por lo que TERRA también tiene características típicas de los RNAs transcritos por la misma, la presencia de una caperuza (m7G) en el extremo 5´y el 7% de los TERRA totales están poliadenilados [152,154] (Figura 7).

4.2. Regulación de TERRA

Un hecho que pone de manifiesto la importancia potencial que puede tener TERRA, es el hecho de su expresión y abundancia están altamente controladas. Una de las primeras observaciones sobre esto, es que TERRA está regulado a lo largo del ciclo celular, alcanzando su máxima abundancia en la fase G1 y decreciendo a lo largo de la fase S [154]. TERRA presenta una regulación pre-transcripcional mediada por distintos factors de transcripción, siendo uno de los más importantes y de los primeros descritos CTCF [155,156]. El estado epigenético del telómero, también influye en este tipo de regulación, por ejemplo, tanto la metilación del DNA subtelomérico, como la presencia de las marcas de heterocromatina H3k9me3, H4K20me3 y HP1 reprimen la expresión de TERRA, mientras que la marca de la eucromatina H3k9ac, la promueve [152,157–159]. Además de la heterocromatinización del telómero, la presencia de algunas shelterinas en el mismo también regula los niveles de TERRA en ciertos tipos celulares, como ocurre con el caso TRF2, cuya eliminación incrementa los niveles de TERRA [160]. Esto apunta al hecho de que situaciones donde la estructura del telómero cambia, podría afectar a la expresión de TERRA y de hecho, durante el acortamiento telomérico y la reprogramación celular, donde la cromatina telomérica se eucromatiniza, la expresión de TERRA aumenta [10,159]. Además, algunos tipos celulares tradicionalmente asociados a una cromatina telomérica menos heterocromanizada, presentan niveles de TERRA especialmente altos, como es el caso de las ESC o las células con actividad ALT [10,161–163]. Una vez transcrito, la estabilidad de las moléculas de TERRA, también está regulada. Por ejemplo, los transcritos de TERRA poliadenilados presentan una vida media mayor que los que no lo están [152,154] y esto parece estar mediado por la maquinaria del *nonsense-mediated decay* (NMD) [151,164]. Finalmente, la interacción con distintas proteínas de unión a RNA, también regulan la estabilidad de TERRA, siendo uno de los ejemplos más importantes el de la familia de las hnRNPs [165].

4.3. Funciones de TERRA

Desde su descubrimiento muchas funciones tanto teloméricas como no teloméricas y con posibles mecanismos independientes entre ellas se han asociado a TERRA, estando la mayoría de ellas basada en ensayos *in vitro*, por la ausencia de modelos genéticos KO para TERRA. Una de esas primeras funciones fue la del mantenimiento de la longitud telomérica. Con respecto al mantenimiento de la longitud telomérica mediado por la telomerasa, distintas evidencias han sugerido roles opuestos de TERRA. *In vitro* se ha demostrado que TERRA es un potente inhibidor de la telomerasa [152,166] y de hecho el silenciamiento de TERRA en ESCs, incrementa la elongación telomérica [167]. En oposición a esto células humanas con altos niveles de TERRA son capaces de elongar los telómeros por telomerasa de forma adecuada [158]. Además, se ha visto que la acumulación de TERRA en telómeros cortos en levadura favorecía el reclutamiento de la telomerasa [168]. Con respecto al ALT, telómeros con más unión de TERRA

favorecen la recombinación de los mismos y su elongación [169,170] y este mecanismo estará asociado a la formación de R-loops, (unas estructuras híbridas que se forma cuando por homología de bases, una cadena de RNA interacciona con una hebra de DNA, desplazando a la hebra contraria), aunque un exceso de estos R-loops, crearía demasiada recombinación y la pérdida de secuencias teloméricas [163]. Otro mecanismo por el que TERRA podría mediar la longitud telomérica sería a través de su papel en la correcta replicación del telómero [171].

Otra función importante de TERRA sería el mantenimiento de la estabilidad del telómero, esta observación se realizó mediante el silenciamiento de TERRA y el hecho de que este silenciamiento causa un incremento del daño telomérico [172,173]. Además, en ESCs, este mecanismo parece estar asociado al desplazamiento de ATRX mediado por TERRA en el telómero [167]. Pero TERRA también podría tener la función contraria y favorecer la inestabilidad telomérica por la formación de R-loops en el mismo [174,175].

Finalmente, el último rol asociado a TERRA en el mantenimiento del telómero, ha sido el de favorecer la correcta heterocromatinización del mismo. En mamífero TERRA es capaz de interaccionar con marcas de heterocromatina telomérica como H3K9me3 y HP1 [173] y con Suv39h1 [160]. El silenciamiento de TERRA además favorece la pérdida de la heterocromatina telomérica [173] y el incremento de su expresión a través del aumento del daño telomérico coincide con una ganancia de H3K9me3 en el telómero [160]. Además, esta función epigenética podría no estar limitada al telómero y se ha visto que TERRA además podría regular la expresión génica. Al aumentar la cantidad de repeticiones teloméricas de RNA en células humanas de cáncer, se genera un cambio general de la expresión génica y esto podría estar causado por la formación de G-cuádruplex entre TERRA y distintos promotores [176]. Además, se ha identificado que TERRA es capaz de unirse a distintas zonas extratelomércas, asociadas a la marca H3K9me3 y H3K27me3 (una marca de heterocromatina sintetizada de forma exclusiva por el complejo PRC2) y su silenciamiento cambia la expresión génica global en ESC, en un mecanismo probablemente mediado por ATRX [167].

4.4. Origen cromosómico de la transcripción telomérica

Aunque se han descrito una gran multitud de funciones para TERRA, estas están principalmente basadas en estudios *in vitro*. Esto es debido a que en los organismos donde se ha identificado la presencia de TERRA, la caracterización se ha hecho principalmente a través de la identificación de su secuencia telomérica de RNA, pero el origen genómico de TERRA y la secuencia subtelomérica que lo compone no está claro. Un aspecto a debate es si todos o muchos telómeros se transcriben, si lo hacen con la misma eficiencia o son solo unos pocos los que dan origen a la expresión de TERRA. Además, esta localización genómica no parece estar conservada entre especies.

En *Saccharomyces cerevisiae*, originalmente se describió que se transcribían de unos telómeros que contenían el llamado elemento Y´, que es un secuencia repetitiva, conservada y presente en aproximadamente el 50% de los telómeros de este organismo, aunque también de

los telómeros que no lo contenían. El extremo 5´de TERRA que contiene el elemento Y´ es relativamente homogéneo, lo que fundamenta la idea de que hay un sitio de inicio de la transcripción definido [147]. Más tarde se caracterizaron siete orígenes transcripcionales concretos [177] y por último se han encontrado evidencias de que en este organismo TERRA no se transcribe a lo vez de todos sus posibles orígenes, sino de un solo locus cada vez [168].

En ratón, se ha caracterizado en iPSCs y en MEFs dos locus distintos, el telómero 18q y el 9q, siendo este primero el origen principal de los TERRA totales. Además, los transcritos del cromosoma 18q pueden unirse a todos los telómeros en *trans*, y su silenciamiento causa un incremento general del daño telomérico, siendo la primera prueba de que los TERRA de un solo locus pueden ejercer funciones en todos los telómeros [172]. De una forma similar en ESCs, se describió que el origen principal de TERRA era la región pseudoautosómica de los cromosomas X e Y [178]. Posteriormente al mapearse los sitios de unión extratelomericos por CHIRT-seq en ESCs, ambos estudios se vieron parcialmente confirmados, al ser estos tres subtelómeros los más altamente unidos por TERRA [167].

En humano en cambio los primeros promotores de TERRA caracterizados fueron promotores basados en islas CpG, caracterizadas por la presencia de las llamadas repeticiones 61-29-37 y localizadas directamente antes del inicio transcripcional de TERRA, encontrándose estos en 18 posibles subtelómeros [157]. Posteriormente usándose técnicas de *pull-down* de RNA, seguidos de RNA-seq, se caracterizaron 18 posibles orígenes de TERRA en células HeLa. Estos 18 presentaban dos tipos distintos, aquellos donde la transcripción empezaba entre 5-10 Kb del telómero y aquellos donde empezaba a 1Kb o menos [160]. Pero más allá de la secuenciación, en este trabajo, no se aportó ninguna evidencia de que los transcritos derivados del mismo, presentaran características propias de TERRA. El primer gran objetivo de esta tesis, ha sido caracterizar si alguno de estos loci presentaba características propias de TERRA, para poder identificar un locus claro en humanos. Esto tiene el objetivo de generar células humanas KO para TERRA y usarlas para esclarecer la función de TERRA en la biología del telómero.

5. Telómeros y pluripotencia

5.1. Introducción, ESCs y iPSCs

Las células madre embrionarias o *Embryonic Stem Cells* (ESCs), son células pluripotentes con capacidad de autoregenerarse y que se pueden diferenciar en cualquiera de las tres capas germinales (ectodermo, mesodermo o endodermo). Estas células derivan del embrioblasto (*Inner cell mass*, ICM) de los blastocitos, un estado temprano del embrión antes de su implantación, y tanto las derivadas de ratón como las de humano se pueden cultivar *in vitro* [179–181]. Estas células en cultivo normalmente están en el estado denominado *primed*, que representaría el estado de las células del epiblasto tras su implantación, pero se pueden llevar a un estado preimplancion también denomina *naïve* tras la adición de 2i, un conjunto de dos inhibidores (para MEK y GSK3), mostrando estos dos estados unas características epigenéticas y funcionales muy distintas [182,183].

Durante el desarrollo, estas células se diferencian y pierden gradualmente su capacidad de autoregeneración y su pluripotencia. Sin embargo, las células ya diferenciadas, pueden devolverse a un estado más pluripotente, mediante un mecanismo denominado reprogramación nuclear, que conlleva grandes cambios en la estructura de la cromatina y la expresión génica [184–187]. La reprogramación nuclear puede conseguirse de tres maneras, la primera de ellas es la transferencia nuclear de células somáticas, que consiste en la inserción de un núcleo de una célula somática a un ovocito enucleado de la misma especie [188]. La segunda es la fusión de una célula diferenciada con una ESC [189]. Y finalmente el método más reciente y más prometedor es el denominado como reprogramación celular y consiste en la reprogramación directa de una célula diferenciada a través de la expresión ectópica de unos factores transcripcionales (denominados factores de Yamanaka), dando origen a unas células con características similares a las de las ESCs, denominadas células madre pluripotentes inducidas o *Induced Pluripotent Stem Cells* (iPS or iPSCs) [190].

Las iPSCs al igual que las ESCs, presenta la capacidad de autoregenerarse y diferenciarse en cualquiera de las tres capas germinales, además de poder contribuir a la generación de ratones quiméricos y a la línea germinal de los mismos [190]. Las iPSCs tanto murinas como humanas, presentan características, morfológicas, moléculas, de crecimiento y de diferenciación muy similares a las de las ESCs, incluyendo patrones similares de expresión génica y epigenéticos [191,192].

5.2. Importancia de la telomerasa en reprogramación y pluripotencia

Tanto las ESCs como las iPSCs, se pueden dividir en cultivo de forma infinita, esta capacidad ilimitada de proliferación, requiere de una gran integridad genómica y por tanto de la elongación del telómero y de su correcto mantenimiento. Este fenómeno se ha explorado ampliamente poniéndose de manifiesto que el telómero es un componente esencial en el mantenimiento de estas células y en el proceso de reprogramación.

Una de las mayores pruebas de esto es que tanto las iPSCs como las ESCs muestran altos niveles de actividad telomerasa [190,193]. Además, los factores de Yamanaka son importantes para la regulación de la expresión de la telomerasa durante la reprogramación. C-MYC, activa la expresión de TERT [194–196], pero su expresión ectópica no es esencial para el proceso de reprogramación o la activación de la telomerasa [10,197]. KLF4 activa la expresión de la telomerasa al unirse a su promotor tanto en cancer como en ESCs [198] y también mediante el reclutamiento de la β-catenina [199]. De la misma forma, la unión de OCT4 y NANOG a dicho promotor, ha sido descrito en iPSCs humanas y ESCs murinas [200]. Sin embargo, la expresión murina de TERT solo se incrementa moderadamente durante el proceso de reprogramación, mientras que en células humanas TERT aumenta su expresión de forma drástica [201].

La elongación de los telómeros mediada por telomerasa en las iPSCs ocurre principalmente tras el proceso de reprogramación [10,200]. Las iPSCs de ratón muestran una

elongación telomérica moderada a pases cortos, con respecto a los MEFs parentales, pero esta longitud va aumentando progresivamente con los pases, hasta alcanzar una longitud similar a la de las ESCs [10]. Aunque el principal mecanismo para la elongación telomérica en ESCs y en iPSCs durante la reprogramación está mediado por la actividad telomerasa, el ALT también se ha visto implicado en este proceso. De hecho, se observan más T-SCE en ESCs y en iPSCs que en MEFs [10] y este mecanismo puede ser fundamental en el mantenimiento de la longitud de telomérica en iPSCs y ESCs en cultivo, estando asociado a la expresión de Zscan4 [73,74]. Además de en la longitud telomérica, como ya se mencionó en el apartado 3.4, durante la reprogramación, las IPSCs sufren un cambio en su cromatina telomérica, pareciéndose esta más a las de las ESCs, que a la de los MEFs, es decir, presentan una pérdida de marcas de heterocromatina y una ganancia de marcas de eucromatina telomérica [10,79].

Pero las demostraciones más claras de la relevancia de la elongación telomérica durante la reprogramación y para el mantenimiento de la pluripotencia, se han conseguido a través de la sobreexpresión o eliminación de componente de la telomerasa. Tanto las ESCs como las iPSCs humanas y de ratón con telómeros cortos, debido a la eliminación de TERT o TERC muestras problemas a la hora de la diferenciación y una pérdida de la pluripotencia [10,202,203]. En contraste con esto, la sobreexpresión de TERT en ESCs, confiere a estas células ventaja proliferativa, protección contra el daño oxidativo y favorece la correcta diferenciación [204,205]. Además, la eliminación de la telomerasa, genera una gran disminución en la eficiancia de reprogramación, tanto en la reprogramación *in vitro*, como *in vivo* [10,206]. En el caso contrario, se ha descrito que la sobreexpresión de TERT en queratinocitos humanos, favorece la eficiencia de la reprogramación [207].

5.3. TRF1, pluripotencia y reprogramación

Pero no solo la telomerasa es importante para la pluripotencia y la correcta generación de iPSCs, otros componentes del telómero, también lo son, como es el caso de las shelterinas. Algunos ejemplos de esto son que TPP1 es necesario para la elongación del telómero, al facilitar el reclutamiento de la telomerasa durante la reprogramación y que se haya visto que TRF1 y TRF2 son marcadores de células madre y que están sobreexpresados en ESCs y en iPSCs [208,209].

En particular, se ha visto que TRF1 podría tener una gran importancia en estos estos fenómenos. Una de las primeras pruebas de esto, es que los ratones KO para TRF1 presentan letalidad embrionaria, y esta ocurre en eventos muy tempranos de desarrollo, en concreto en el estado de blastocito. Además, estos blastocitos no presentan ningún defecto telomérico, ni en la longitud telomérica ni en el daño telomérico [210]. TRF1 es uno de los genes que presentan más sobreexpresión en ESCs y iPSCs y durante la reprogramación *in vitro* e *in vivo*) [206,209,211,212]. Y esta alta expresión está asociada a eventos tempranos de la generación de las ESCs, coincidiendo con una alta expresión de SOX2 y OCT4 [213]. Además, OCT4 se une directamente al promotor de TRF1 y modula su expresión en células pluripotentes [209]. Por último, sin TRF1, el

proceso de reprogramación no puede ocurrir *in vitro* [209], y su silenciamiento a través de inhibidores, baja la eficiencia de reprogramación *in vivo* [206].

Todos estos datos ponen de manifiesto la gran importancia de TRF1 para el correcto mantenimiento de la pluripotencia y el proceso de reprogramación, y dejan ver que esto podría estar medido por un mecanismo distinto a la de la protección telomérica. Entender si este mecanismo existe y la caracterización del mismo, es el segundo gran objetivo de esta tesis doctoral.

Objetivos

Objetivos

Estudiar el papel de los *Telomeric repeat–containing RNA* (TERRA) en la biología del telómero. Con este propósito nuestros objetivos son:

-Encontrar *locus* que transcriban TERRA en células humanas.

-Caracterizar el promotor de estos *locus.*

-Generar células humanas TERRA KO.

-Estudiar el papel de TERRA en la estabilidad y longitud telomérica.

-Estudiar el papel de TERRA en la epigenética del telómero.

Estudiar el papel de la proteína telomérica TRF1 en el mantenimiento de la pluripotencia. Con este propósito nuestros objetivos son:

-Caracterizar los cambios transcripcionales asociados al silenciamiento de TRF1 en IPSCs.

-Confirmar que estos cambios no son debidos a un incremento en el daño telomérico.

-Identificar los cambios epigenéticos que desencadenan de los cambios de expresión.

-Estudiar el mecanismo que genera estos cambios epigenéticos y de expresión.

Artículos

Artículo 1

Telomeric RNAs are essential to maintain telomeres

Artículo 1: Telomeric RNAs are essential to maintain telomeres.

Autores: Juan José Montero*, Isabel López-Silanes*, Osvaldo Graña y María A. Blasco
(* Estos autores han contribuido de igual manera al manuscrito).

Publicado en Nature Communications. 17 de agosto de 2016. 7, número de artículo: 12534

Resumen:

Los telómeros son estructuras nucleoprotéicas altamente heterocromatizadas situadas al final de los cromosomas eucariotas y que protegen a los mismos de ser reconocidos como daño en el DNA. A pesar de su alta heterocromatinización, los telómeros son capaces de transcribirse en unas lncRNAs conocidos como *Telomeric repeat-containing RNA* (TERRA). Desde su descubrimiento muchas funciones han sido asignadas a TERRA, pero a pesar de ello, pocas se han podido validar en modelos celulares debido a la falta de modelos *knock-out* (KO) para TERRA. Esta falta de modelos KO está causada porque se desconocía los *loci* a partir de los cuales TERRA se transcribía.

Para tratar de solventar este problema, en este estudio nos enfocamos en confirmar la presencia de características propias de TERRA en los 18 *loci* previamente propuestos como inicios transcripcionales de TERRA en humanos. Sorprendentemente vimos que el 83,3% de estos locus se encontraban en subtelómeros estructuralmente conservados. Cuando evaluamos por distintas metodologías (RNA-FISH y Northern) si los transcritos procedentes de estas regiones presentaban características típicas de TERRA, no pudimos encontrar ninguna. En cambio, los transcritos procedentes del subtelómero 20q y Xp si presentaban características propias de TERRA. Posteriormente generamos líneas celulares humanas de cáncer U2OS que mantenían los telómeros por ALT, en las que eliminamos en homocigosis las regiones 20q o Xp mediante la tecnología CRISPR. El KO del *locus* Xp no afectaba los niveles de TERRA totales. Sin embargo, el KO de la región 20q disminuye los niveles de TERRA drásticamente, indicando que esta región es un *bona fide locus* de TERRA.

Usando las líneas células U2OS 20q-TERRA KO, hemos estudiamos el efecto de la eliminación de TERRA en el mantenimiento del telómero. En concreto observamos una disminución de la longitud telomérica, un aumento del daño en el DNA total y el DNA telomérico. Además, vimos un aumento de la inestabilidad génica en las células 20q-TERRA KO con respecto a las WT.

En conclusión, en este estudio hemos generado por primera vez en cualquier organismo, un KO para un *locus* de TERRA, lo que nos ha permitido describir el papel crítico de TERRA en el mantenimiento de la estabilidad y longitud telomérica.

Artículo 1

Contribución Personal: He participado en el diseño del estudio, realizado y analizado la mayoría de los experimentos (diseño de sondas, localización del *locus* de TERRA, realización de ensayos luciferasa, puesta a punto del sistema CRISPR, generación y confirmación de las líneas celulares TERRA KO y análisis de la longitud y del daño telomérico). Finalmente, he contribuido en la discusión de los resultados y en la preparación del artículo al elaborar las figuras.

nature COMMUNICATIONS

ARTICLE

Received 19 Mar 2016 | Accepted 11 Jul 2016 | Published 17 Aug 2016

DOI: 10.1038/ncomms12534 OPEN

Telomeric RNAs are essential to maintain telomeres

Juan José Montero[1,*], Isabel López de Silanes[1,*], Osvaldo Graña[2] & Maria A. Blasco[1]

Telomeres are transcribed generating long non-coding RNAs known as TERRA. Deciphering the role of TERRA has been one of the unsolved issues of telomere biology in the past decade. This has been, in part, due to lack of knowledge on the TERRA loci, thus preventing functional genetic studies. Here, we describe that long non-coding RNAs with TERRA features are transcribed from the human 20q and Xp subtelomeres. Deletion of the 20q locus by using the CRISPR-Cas9 technology causes a dramatic decrease in TERRA levels, while deletion of the Xp locus does not result in decreased TERRA levels. Strikingly, 20q-TERRA ablation leads to dramatic loss of telomere sequences and the induction of a massive DNA damage response. These findings identify chromosome 20q as a main TERRA locus in human cells and represent the first demonstration in any organism of the essential role of TERRA in the maintenance of telomeres.

[1] Telomeres and Telomerase Group, Molecular Oncology Programme, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain.
[2] Bioinformatics Group, Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain.
* These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.A.B. (email: mblasco@cnio.es).

TERRA transcripts are nuclear long non-coding RNAs that are transcribed from the subtelomere towards the telomere[1,2]. They are transcribed by RNA polymerase II, giving rise to transcripts that contain UUAGGG-repeats, being the presence of this repeat their main features. They are also heterogeneous in size (0.2–10 kb in humans and mice) as indicated by the smear detected in TERRA northern blots[1,2]. On the other hand, detection of TERRA by RNA-fluorescence *in situ* hybridization (FISH) renders a clear spotted pattern in the nucleus and the number of spots varies in different cell types. Approximately 30% of these spots co-localize with telomeres[1,3].

Lack of TERRA's subtelomeric sequence information has been an important shortcoming to understand the role of TERRA, as most the functional studies published to date are based on the use of the UUAGGG-tract to detect TERRA. Thus, TERRAs have been implicated in telomere protection[3–5], heterochromatin formation[4], telomere replication by sequestering hnRNPA1 from telomeres to allow RPA-to Pot1 switching, a process that it is altered in alternative lengthening of telomeres (ALT) cells lacking *ATRX*[6,7], and in telomere elongation by homologous recombination through the formation of DNA-TERRA hybrids[8,9]. TERRAs are also proposed to bind the telomerase core components, TERC and TERT, but their role in telomerase function is unclear. On one hand, TERRA can inhibit telomerase activity *in vitro*[2,10]. On the other hand, in yeast, TERRAs are induced at short telomeres and form TERRA-TERC RNA clusters that co-localize with the telomere of origin during S phase, suggesting that TERRA may play a role in the spatial organization of telomerase activity at telomeres[11].

In yeast, several TERRA transcription start sites have been identified at different chromosome ends. In the mouse, by using a genome wide RNA-seq approach, we recently identified that transcripts showing bona fide TERRA features mainly arise from the chromosome 18 subtelomere, and to a lesser extent from the subtelomere of chromosome 9 (ref. 3). Furthermore, we found that chromosome 18-TERRA associate in *trans* with the remaining telomeres and are important for telomere protection[3]. TERRA loci in human cells have been more elusive. First, a putative TERRA promoter was proposed to consist of a 61-29-37 repeat (a conserved region that contains three different repetitive DNA tracts of 61, 29 and 37-bp), which is present at 20 different chromosomes[12]. However, transcript sizes and their regulation were not identical to those of TERRA, suggesting that they may constitute a fraction of TERRA molecules[12]. More recently, a similar RNA-seq approach to that used by us in mice, identified 10 distinct human chromosome ends where putative TERRA transcription started as far as 5–10 kb away from the telomere, as well as eight additional chromosome ends where transcription started in close proximity to the telomere[13]. The authors, however, did not address whether these transcripts showed TERRA features, defined as the presence of UUAGGG-repeats within their sequence, their heterogeneity in size and the nuclear spotted pattern in which some spots co-localize with the telomere.

Here, we study whether these human transcripts have the above-mentioned TERRA features with the final goal of functionally deleting them to unveil the role of TERRA. We found that, out of the 18 proposed TERRA loci in humans, only transcripts arising from 20q and Xp loci have TERRA features. We then used the CRISPR-Cas9 technology to genetically delete these two potential TERRA loci in humans to study whether they are indeed the origin of TERRA. Only deletion of the 20q locus caused a dramatic decrease in TERRA levels, while deletion of the Xp locus did not result in decreased TERRA levels. Importantly, deletion of the 20q-TERRA locus leads to telomere shortening and telomere uncapping as indicated by increased telomere damage foci or TIFs. These unprecedented findings demonstrate that TERRA transcripts are essential for the maintenance of a functional telomere cap.

Results

Identification of the human TERRA locus. A number of subtelomeric transcripts identified in samples that underwent UUAGGG-transcript enrichment have been recently proposed to represent the human TERRA transcriptome, although further confirmation that these transcripts had TERRA features (see the first section) was lacking[13]. These RNAs arose from 10 distinct chromosome ends where transcription started as far as 5–10 kb upstream of the subtelomere-telomere boundary (referred here as '10-kb TERRA') and from 8 chromosome ends in which transcription started in close proximity to the telomere (referred here as '1-kb TERRA')[13].

Here, we set out to study whether these subtelomeric transcripts have indeed TERRA features (see the first section), and if so, to generate functional knock-outs by using the CRISPR-Cas9 technology. With this approach, we aimed to unravel the role of TERRA in human cells.

First, we studied the genetic structure of the different identified sequences by aligning them against two different human assemblies, the UCSC human browser GRch37/hg19 and the subtelomeric specific assembly[14]. Bioinformatic analysis revealed that 83.3% of these transcripts arise from DNA regions that are structurally conserved. Thus, the '10-kb TERRA' loci contain three different conserved DNA elements, a repetitive region known as TAR1 (UCSC RepeatMasker track; Smit AFA, Hubley R, Green P. RepeatMasker Open-3.0, www.Repeatmasker.Org. 1996–2010) and two additional conserved DNA regions which encode for two long non-coding RNA families of pseudogenes derived from the primate genes WASH and DDX11 (refs 15,16), respectively (Fig. 1a). From now on, we will refer to these three conserved DNA regions as TAR1, WASH and DDX regions/loci, and to the transcripts arising from these regions as TAR1, WASH and DDX transcripts. The '1-kb TERRA' contained only the TAR1 repetitive region (Fig. 1a). Surprisingly, the DDX region is transcribed in the opposite direction from what expected for a TERRA transcript[15,16] (TERRA is transcribed from the subtelomere towards the telomere)[1], thus suggesting that they are not likely to correspond to TERRA (Fig. 1b). Thus, transcripts arising from the subtelomeres of chromosomes 1p, 9p, 12p, 15q, 16p, 19p, Xq are unlikely to be bona fide TERRAs.

Nevertheless, we designed RNA-FISH and northern blot probes against the transcripts arising from each of the DNA structurally conserved regions, TAR1, DDX and WASH (Fig. 1c) and compared their signals to that of a probe detecting the TERRA's-UUAGGG-tract, as it is the TERRA main feature (see the first section). Sequence degeneration at subtelomeres allowed us to simultaneously detect transcripts arising from various subtelomeres (see probe specificity in Supplementary Tables 1–4). In double RNA-FISH experiments, we failed to see co-localization of sense (S) and antisense (AS) probes against TAR1, DDX and WASH RNAs with transcripts detected with the TERRA's-UUAGGG probe (Supplementary Fig. 1A). Sense and antisense probes were used since TAR1, DDX and WASH transcripts are not transcribed with the same orientation. Moreover, probes against TAR1 transcripts rendered a strong cytoplasmic signal, in contrast to the known nuclear localization of TERRA transcripts (Supplementary Fig. 1A). Of note, in the case of the DDX locus, whose transcripts are transcribed in the opposite direction than what is expected for a TERRA transcript, we failed to see co-localization with TERRA when using either the sense or the antisense DDX probes. As control, RNAse treatment erased all

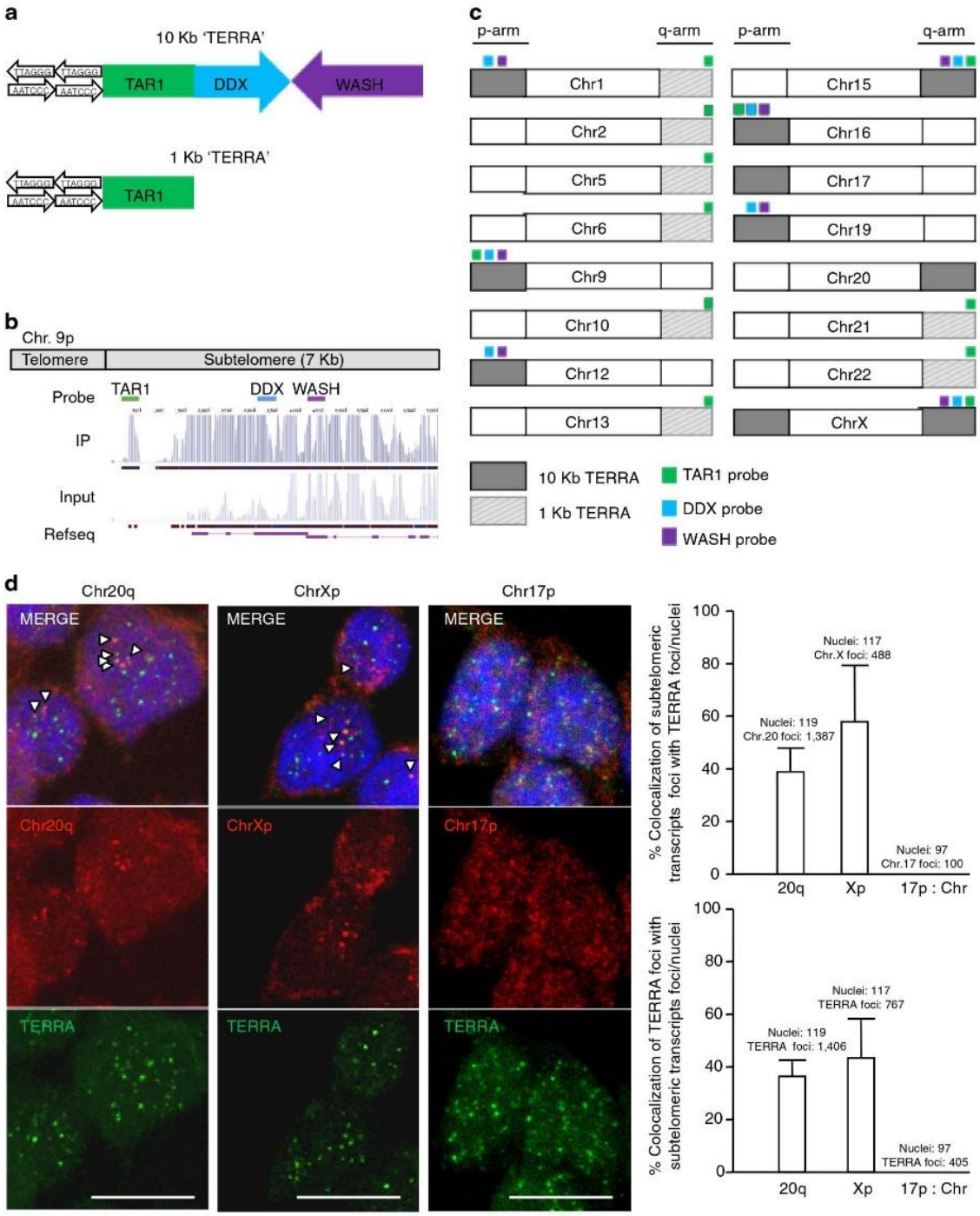
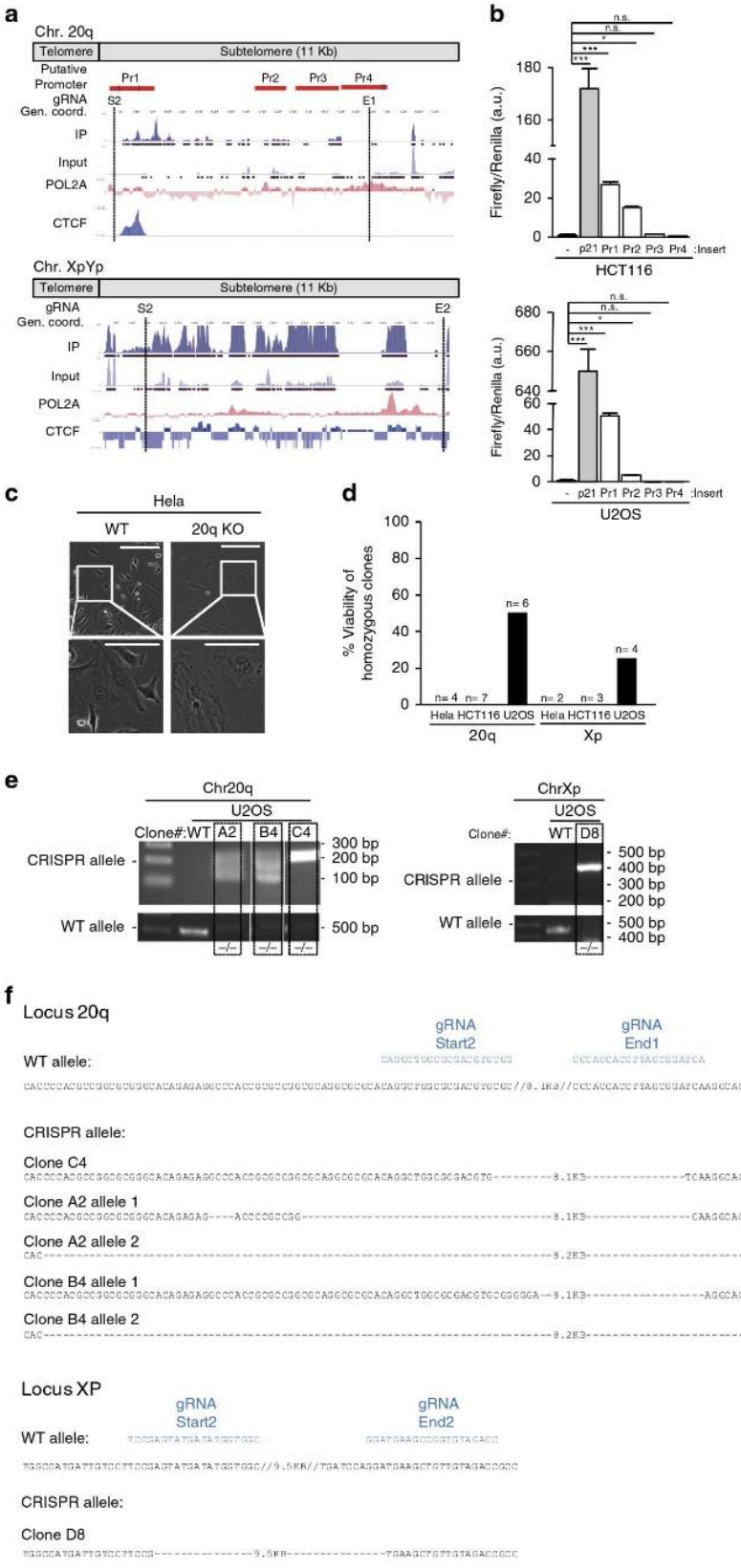


Figure 1 | Transcripts arising from the subtelomere of chromosomes 20q and Xp co-localize with TERRA. (**a**) Schematic of the three motifs in the structurally conserved TERRA loci proposed in Porro *et al.*,[13] the repetitive region TAR1 and the two lncRNA families, DDX11L and WASH. (**b**) Example of the position of the probes against the structurally conserved motifs in the subtelomere of chromosome 9p. The alignment of the RNA-seq reads obtained in a TERRA-IP compared with the input[13] and an annotated Ref seq in this region is also shown. (**c**) Schematic of the position of TERRA loci proposed in Porro *et al.*,[13] within all chromosomes. The 10 kb TERRA is coloured in grey and the 1-kb TERRA with stripes. The position of the probes against TAR1, DDX11L and WASH is also indicated in different colours. (**d**) Confocal microscopy images of double RNA-FISH using probes targeting either subtelomere 20q, Xp or 17p transcripts (red) and TERRA's telomeric tract (green). The representative images are the merge of four individual confocal images. Co-localization events were only counted as positive when detected in the individual confocal images. Arrowheads indicate real co-localization events detected in the individual confocal images. (Top graph) The percentage of co-localization of either subtelomere 20q foci or subtelomere Xp foci with TERRA foci with respect the total number of 20q or Xp foci is represented (mean ± s.d., n = number of nuclei). Total number of nuclei and foci are also indicated. (Bottom graph) The percentage of co-localization of TERRA foci with either subtelomere 20q foci or subtelomere Xp foci with respect the total number of TERRA foci is also represented (mean ± s.d., n = number of nuclei). Total number of nuclei and foci are also indicated. Scale bar, 10 μm.

these different RNA-FISH signals confirming the specificity of the different probes to defect RNA and not DNA (Supplementary Fig. 1B). By using the northern blot analysis, we confirmed that transcripts arising from the TAR1, DDX and WASH loci did not show the expected TERRA pattern (Supplementary Fig. 1C). Furthermore, we did not detect any specific signal from the TAR1 probe, only the unspecific cross-hybridization with the 18S and 28S rRNAs (Supplementary Fig. 1C).

Next, we took advantage of the known effect of DNA methylation on regulating human TERRA levels to evaluate



TERRA features in WASH, TAR1 and DDX transcripts. In particular, DNA demethylating agents are known to increase human TERRA levels[12]. To this end, we studied TERRA levels in the HCT116, Hela and U2OS human cell lines before and after treatment with the demethylating agent 5′ azacytidine (Aza), as well as in HCT116 cells double deficient for DNMT1 and DNMT3b (double knockout, DKO cells) as a positive control[17]. In marked contrast to TERRA transcripts, which were upregulated on Aza treatment, we failed to see induction of DDX, WASH and TAR1 transcripts on Aza treatment (Supplementary Fig. 1C,D), indicating that they do not show the expected TERRA behaviour[12].

As none of the transcripts arising from the DNA structurally conserved loci TAR1, DDX and WASH showed features of TERRA (see the first section), we next studied the TERRA properties in the transcripts arising from the remaining non-conserved loci located at chromosomes 17p, 20q and Xp, which also showed 'read enrichment' in the described TERRA-IP[13]. Interestingly, in double RNA-FISH experiments, we found co-localization of transcripts arising from 20q and Xp subtelomeres with TERRA transcripts (detected with a probe against TERRA's-UUAGGG), suggesting that they could correspond to genuine TERRA transcripts (Fig. 1d). In particular, ~40% of signal spots from the 20q locus co-localized with TERRA signals, reaching 60% in the case of the Xp locus (Fig. 1d, top graph). From the total TERRA spots, 40% co-localized with spots from either 20q or Xp loci (Fig. 1d, bottom graph). Transcripts arising from the 17p locus, however, did not show co-localization with TERRA transcripts (Fig. 1d), thus ruling out that they contribute to TERRA. Note that the co-localizations were counted as positive only when detected in the individual confocal layers (see graphs). Apparent co-localizations between Chr17-RNAs and TERRA in the representative images of Fig. 1d were not counted, as they are the result of a visual effect of image superposition of individual confocal layers. Real co-localization events for Chr20q and Xp-RNAs with TERRA are indicated with arrowheads and were detected in individual confocal layers (Fig. 1d). As control, RNAse treatment also erased these RNA-FISH signals, thus confirming the specificity of the different probes (Supplementary Fig. 1B).

Table 1 | Number and percentage of clones obtained for each genotype on deletion of the 20q and Xp loci with the CRISPR-Cas9 system.

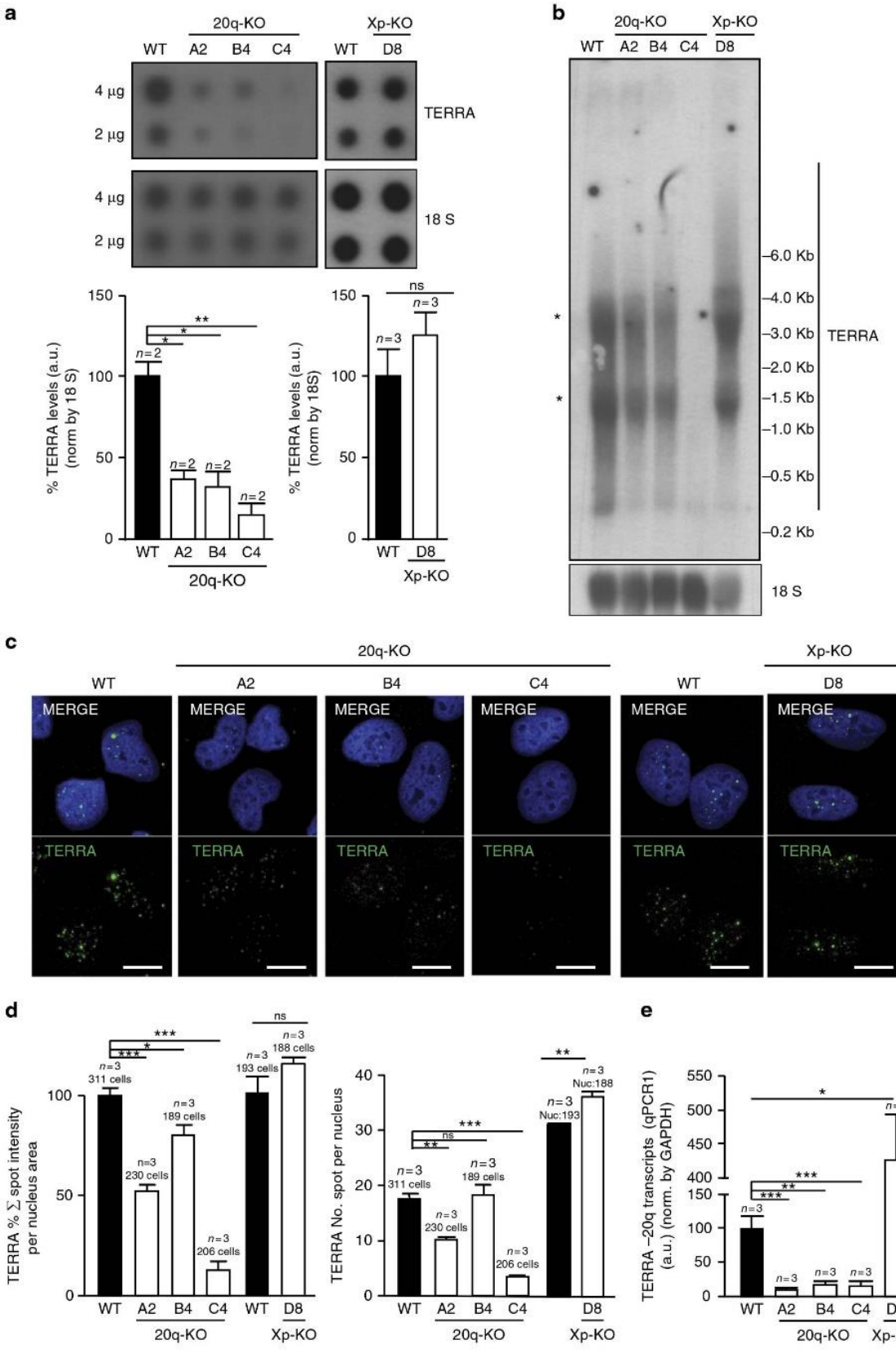
	No. clones	No. +/+ (%)	No. −/+ (%)	No. −/− (%)
20q				
U2OS	55	7 (12.7)	11 (20)	37 (67.3)
HeLa	63	25 (39.8)	34 (53.9)	4 (6.3)
HCT116	47	37 (78)	3 (6)	7 (16)
XpYp				
U2OS	70	50 (71.4)	16 (22.9)	4 (5.70)
HeLa	49	42 (86.1)	5 (9.8)	2 (4.10)
HCT116	32	22 (68.7)	0 (0)	10 (31.3)

Transcripts arising from the 20q locus are bona fide TERRA. To further demonstrate that transcripts arising from 20q and Xp loci are genuine TERRA transcripts, we set to genetically delete the loci from which they are transcribed by using the CRISPR-Cas9 system in the HCT116, Hela and U2OS human cancer cell lines, as well as in the non-transformed IMR90 human fibroblast cell line. First, we performed a detailed transcriptional analysis of the regions to be deleted (Fig. 2a). Previously reported chromatin immunoprecipitation (ChIP)-seq data showed the presence of active RNA polymerase II and CTCF binding sites at both the 20q and Xp loci[14]. Thus, we assessed the presence of promoter activity in these regions by cloning different fragments upstream of the RNA-seq read enriched regions into promoter-free luciferase reporter vectors (Fig. 2a,b). We found a strong promoter activity for the 20q locus in the vicinity of the telomere (region Pr1) and to a lower extent in a region 4.7-kb away from the telomere (region Pr2) both in HCT116 and U2OS human cell lines (Fig. 2b). The p21 promoter was included as positive control (Fig. 2b). In the case of the Xp locus, we were unable to clone fragments of interest to assess its promoter activity owing to the highly repetitive content of this region. Collectively, these findings confirm strong transcriptional activity within the 20q locus.

Next, in order to delete these putative TERRA loci, we designed guide RNA (gRNAs) for both loci. In particular, we set to delete a 8.1-kb region in the 20q locus, which includes the promoter regions, and a 9.5-kb region in the Xp locus. Of note, these regions were the ones showing enrichment of RNA-seq reads in the previously described TERRA-IP[13] (Fig. 2a). Using the online CRISPR designer tool[18], two different gRNAs showing the lower off-target mutation score were selected for each of the flaking regions (gRNA-start 1 and start 2 for the region closer to the telomere and gRNAs-end 1 and end 2 for the region more distal from the telomere) (Supplementary Table 5). We tested the best combination of gRNAs by transfecting two plasmids, each of them containing the Cas9 and the GPF reporter but a different gRNA each, corresponding to each of the flanking regions (Supplementary Table 6). Two days post transfection, the presence of the deletion was studied by PCR using specific primers (Supplementary Table 7). For all the four human cell lines, we successfully detected different deletions of the 20q and Xp loci except for the Xp locus in HCT116 cells (Supplementary Fig. 2A). Next, we set up the conditions to select single clones with the deletion of interest. To this end, we transfected the most efficient combination of gRNAs for each flanking region. Each gRNA was within a plasmid containing the Cas9 and a green

Figure 2 | Deletion of the 20q-TERRA locus in different human cell lines using the CRISPR-Cas9 system. (**a**) Snapshot depicting, from top to bottom, putative promoter regions (Pr1, Pr2, Pr2 and Pr4), gRNA position, genomic coordinates, RNA-seq from TERRA IP and input[13], and RNA polymerase 2A (POL2A) and CTCF ChIP-seq data. S2, E1, S2 and E2 are the names of the different gRNAs. (**b**) Graph shows the relative fold increase in firefly luciferase activity seen in the pGL3-containing promoter regions (Pr1-4) relative to the empty vector after normalization to renilla activity. (Mean values ± s.e.m., $n=3$ independent experiments). p21 promoter serves as positive control. One-way Anova with the Dunnett's post test was used for statistical analysis (*$P<0.05$ and ***$P<0.001$). (**c**) Representative image of the HeLa clones homozygous for the 20q-TERRA deletion during expansion. Zoom areas are shown. Scale bar, 500 μm and (zoom) 250 μm. (**d**) Percentage of viability of the different homozygous clones for the 20q and Xp deletion on expansion of the three different cell lines. (**e**) Ethidium bromide gels showing the CRISPR allele for the deletion of the 20q and Xp loci detected by PCR in different clones of the U2OS cells. The white strips in the gels indicate the removal of irrelevant samples from that gel. (**f**) Schematic of the sequencing of the CRISPR allele for the deletion of the 20q (clones A2, B4 and C4) and Xp (clone D8) loci compared with the WT allele. Slashes (//) represent omitted DNA sequence. The size of the omitted sequence is shown. Dashes (-) represent the deleted sequence. The size of the deleted sequence is shown. gRNAs are shown in blue.

fluorescent protein (GFP) reporter. Two days post transfection, the top 10% GFP brightest cells were sorted by flow cytometry and seeded into 96-wells plates (1, 3, 5 or 10 cells per well). Two weeks after wells were checked for clonal cell expansion, and only those coming from single clones were genotyped by PCR (Supplementary Table 7). Clonal cell expansion was successful in all cell lines except for the non-transformed IMR90 fibroblasts (see the number and percentage of clones obtained for each



genotype in Table 1). We then set to expand clones homozygous for the 20q and Xp deletions from Hela, HCT116 and U2OS cell lines. During expansion of Hela clones with the 20q-deletion, three out of four homozygous clones developed intense vacuolization and died (Fig. 2c,d). The remaining homozygous clones bearing either the 20q or the Xp deletion from the three cell lines (Hela, HCT116 and U2OS), except for few U2OS clones (see below), either did not grow or were found to be heterozygous for the deletion on the re-genotyping performed after expansion (Fig. 2d; Supplementary Fig.3A). Thus, we were unable to isolate viable clones with deletions in either 20q or Xp in both Hela and HCT116 cell lines, indicative of strong lethality of the deletions (Fig. 2d). Only in the case of U2OS cells, we were able to expand three clones out of six bearing the 20q-deletion (only six clones were expanded from the original 37 KO clones obtained; Table 1) and one clone out of four with the Xp deletion, which were confirmed to be homozygous both by PCR and by Sanger sequencing (clones A2, B4 and C4 for the 20q locus and clone D8 for the Xp) (Fig. 2e,f). The smear observed in the detection of the CRISPR-Cas9 alleles in clones A2 and B4 can be resolved in two distinct bands (Supplementary Fig. 3B) and it is due to two different deletions in each of the alleles of these clones within the 20q locus (see sequences displayed in Fig. 2f). Interestingly, we observed that U2OS cells display higher basal levels of TERRA compared with Hela and HCT116 cells (Supplementary Fig. 1B,C), which may explain the differential effects of TERRA abrogation on cell viability of the three different human cell lines studied.

Next, we checked the expression of total TERRA levels by RNA dot-blot and northern blot in the individual clones deleted for the 20q and Xp loci. Deletion of the 20q locus resulted in a dramatic reduction in total TERRA levels in all three clones isolated from U2OS cells (see clones A2, B4 and C4 deleted for the 20q locus in Fig. 3a,b). In contrast, deletion of the Xp locus did not result in changes in total TERRA expression (Fig. 3a,b; Supplementary Fig. 4). The lack of TERRA downregulation as the result of the Xp deletion, may indicate that this is not a major locus for TERRA production in human cells. Alternatively, this may be related to the fact that the region deleted in the Xp locus was smaller than that deleted in the 20q locus, owing the impossibility to design gRNAs in the vicinity of the Xp telomere due to its highly repetitive nature.

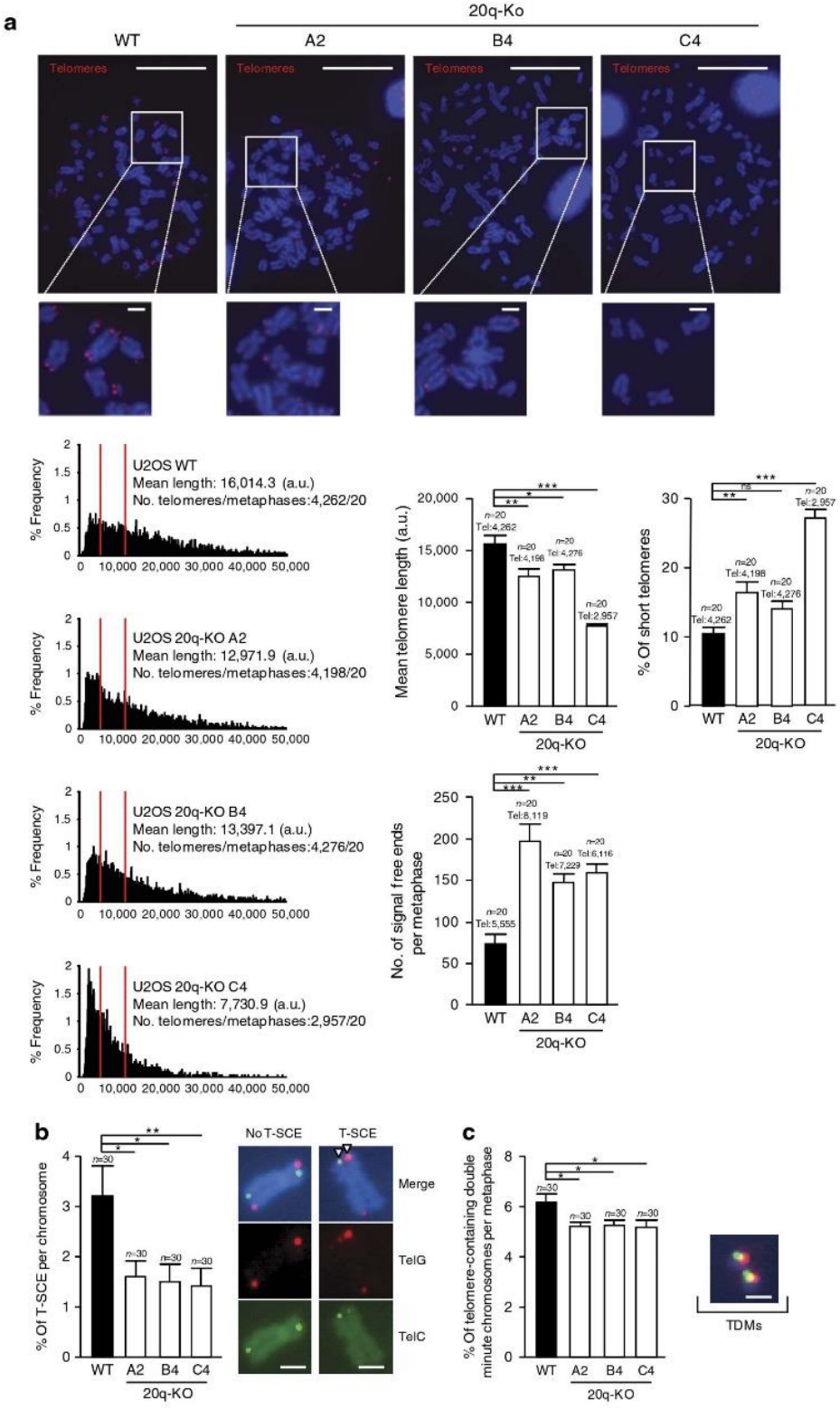
Next, we confirmed TERRA downregulation owing to the deletion of the 20q locus but not of the Xp locus by several other independent techniques. In particular, RNA-FISH using probes that detect the TERRA's-UUAGGG-tract also demonstrated a dramatic decrease of the total TERRA spot intensity per nucleus in the 20q locus deleted clones but not in the Xp locus deleted one when compared with wild-type (WT) cells (in Fig. 3c,d, left graph). This was paralleled by a significant decrease in TERRA spots per nucleus in the 20q locus deleted clones but not in the Xp deleted one (Fig. 3d, right graph). Downregulation of TERRA levels in the 20q-deleted clones but not in the Xp deleted clone, was also confirmed by quantitative PCR (qPCR) (Fig. 3e; Supplementary Table 8).

To ascertain whether the changes in TERRA levels and in the phenotype observed (see below) could be due to the presence of CRISPR-Cas9 off-target mutations, we performed TIDE (Tracking of Indels by DEcomposition) analysis for each of the gRNAs used to generate the 20q-deletion (named Start1 and End1). We selected for study the five regions with the highest off-target mutation score but only the first four could be studied because of the impossibility to amplify the fifth region for each gRNA, probably owing to sequence differences between the U2OS genome and the human reference genome (GRCh38/hg38). The analysis was performed on DNA pools from either WT cells or cells bearing the 20q-deletion, named C3 (from which the clones A2, B4 and C4 were obtained). We did not find off-target mutations for any of the gRNAs except for the gRNA END1/off-target3 (Supplementary Fig. 5A). To verify whether this off-target mutation was present in the 20q-KO clones A2, B4 and C4, we carried out TIDE analysis on DNA from these clones. The 20q-KO clone A2 was free of gRNA END1/off-target3 but the clones B4 and C4 revealed 3–7% presence of off-target mutations (Supplementary Fig. 5B). It is unlikely that this was a real off-target mutation (vs the result of poor quality sequencing) because: (i) if an off-target mutation was present in these clones the percentage of off-targets should be 50 or 100% owing to the clonality of the cells, and (ii) comparison with DNAs from cells untreated with CRISPR revealed a similar percentage of off-target mutations (Supplementary Fig. 5B). Nevertheless, to confirm absence of off-target mutations, we performed T7-endonuclease digestion on DNA PCR amplified from the region of interest in both non-CRISPR-treated cells and from the CRISPR 20q-KO pool and clones (pool C3 and clones A2, B4 and C4). As a positive control, we used two identical DNAs except for a C to G mutation (DNAs were named C and G). As expected, two digestion bands were seen in the C + G digestion but no digested bands were observed neither in the C + C digestion nor in the cells that underwent the CRISPR deletion of the 20q locus (pool C3 and clones A2, B4 and C4) (Supplementary Fig. 5C). This assay confirms the absence of off-target 3 in the 20q-KO clones treated with the END1 gRNA. Note that the sequence alignment for the START2/off-target2 and for the END1/off-target4 regions was poor due to technical problems during sequencing because of the presence of long tracks of C or G within the sequences of these regions (Supplementary Figure 5A). Although we did not find off-targets mutations in the TIDE analysis of these regions, we further confirmed this finding with the T7-endonuclease assay (Supplementary Figure 5C).

Collectively, these data indicate that transcripts arising from the 20q locus are genuine human TERRA transcripts, which account for a large proportion of TERRA levels, while we failed to see TERRA downregulation when deleting the Xp locus. From now on we will refer to the 20q transcripts as '20q-TERRA' transcripts.

Figure 3 | Deletion of the TERRA-20q locus dramatically affects TERRA expression. (**a**) RNA from the U2OS cells WT or from the three 20q-KO clones (A2, B4 and C4) and the Xp-KO (D8) clone was isolated and used for TERRA detection by RNA dot-blot with a probe against the TERRA-UUAGGG-tract; 18S serves as loading control. (Graph) TERRA quantification normalized by 18S (mean values ± s.e.m.). (**b**) Northern blotting using 32P-dCTP-labelled probe against TERRA-UUAGGG-tract in the U2OS cells WT or KO for the 20q or Xp loci. 18S was included as a loading control. *Unspecific band due to cross-hybridization with rRNA 18S and 28S. (**c**) Representative confocal microscopy images of RNA-FISH against TERRA-UUAGGG-tract (green) in the U2OS cells WT and KO for the 20q (clones A2, B4 C4) and the Xp (clone D8) loci. Scale bar, 10 μm. (**d**) The graphs show (left) the quantification of the total spot intensity per nucleus normalized by nucleus area, (right) the total number of spots per nucleus in the three 20q-KO clones and in the Xp-KO clone D8 (mean values ± s.e.m., $n = 3$ independent experiments). (**e**) Detection of the 20q-TERRA transcripts by qPCR (primers were designed in the subtelomeric region). The percentage of enrichment of the 20q-TERRA transcripts in WT and in the 20q-KO clones (clones A2, B4 and C4) and in the Xp-KO (clone D8) normalized by GAPDH is shown. One-way Anova with Dunnett's post test was used for the statistical analysis of the 20q clones and the Student's *t*-test for the Xp clone (*$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$).

20q-TERRA locus deletion leads to loss of telomeric sequence. Identification of 20q as one of the major locus for human TERRA generation allows us to address the role of TERRA in telomere biology. Thus, we set to assess the impact of the 20q-TERRA locus deletion in telomere length and telomere protection. As shown in Fig. 4a, deletion of the 20q-TERRA locus resulted in a dramatic loss of telomeric sequences in all three clones (A2, B4 and C4) as seen by a significant decrease in individual telomere fluorescence intensity as determined by quantitative telomere FISH (Q-FISH) on single telomeres from metaphase spreads (see lack of detectable telomere signals in metaphasic chromosomes in Fig. 4a, and the dramatic change in the distribution of individual telomere signals), as well as by a significant increase in the percentage of signal-free ends per metaphase and in the



percentage of short telomeres as determined by low telomere fluorescence (short telomeres are considered those in the 10% percentile of the total telomere length distribution) (Fig. 4a and graphs). Given that U2OS cells lack telomerase activity and elongate telomeres via a telomerase-independent mechanism that relies on homologous recombination, the so-called ALT[19,20], we set to address whether the dramatic telomere shortening induced by deletion of 20q-TERRA locus was accompanied by decreased recombination events between telomeres. To this end, we measured the rates of telomeric sister chromatid exchanges (T-SCE) by using chromosome orientation FISH or CO-FISH[21]. Although we found T-SCE events in all the 20q-TERRA KO U2OS clones, their frequency was significantly decreased compared with the non-deleted U2OS controls (Fig. 4b). We also observed a significant decrease in telomere-containing double minute chromosomes on 20q-TERRA deletion in all three clones, a chromosomal aberration also associated with ALT (Fig. 4c)[22]. These results are in line with the proposed role of TERRA in telomere recombination[8,9]. However, the extremely short telomeres present in 20q-TERRA deleted cells, even in the presence of low but detectable telomere recombination events, suggests a more general role for TERRA in the maintenance of proper telomere homoeostasis.

Next, we addressed the effects of TERRA deletion on telomere protection and chromosomal instability. First, we assessed telomere protection by determining the presence of either γH2AX or 53BP1 DNA damage foci at telomeres (the so-called telomere induced foci or TIFs). To this end, we performed double immunofluorescence with either γH2AX or 53BP1 antibodies to detected DNA damage foci combined with TRF2 immunofluorescence to detect telomeres. We found a significant increase in total DNA damage as indicated by increased γH2AX and 53BP1 fluorescence levels in all three clones except for the clone B4 in the case of 53BP1 (Fig. 5a, b). Interestingly, in all three clones and using both γH2AX and 53BP1, we found a significant increase in telomere damage foci or TIFs compared with WT controls (Fig. 5c,d). Finally, in agreement with short telomeres and increased telomere damage, phenotypes that were particularly severe in clone C4, we also found significantly increased end-to-end fusions in this clone (Fig. 5e). We further confirmed increased chromosomal instability in the 20q-TERRA KO C4 clone by comparative genome hybridization (CGH). We found dramatic genome reorganizations in the 20q-KO C4 cells compared with WT controls (Fig. 5f). In particular, we found chromosomal gains in chromosomes 7, 9, 14 and 19 along with large chromosomal losses in chromosomes 4, 6, 10, 15 and 21, including the complete loss of one of the arms in chromosomes 6 and 18, in the 20q-KO cells with respect WT cells. Some other smaller losses were also detected in chromosomes 3, 12, 18 and 19 (Fig. 5f). A complete list of the genomic coordinates of the different chromosomal gains and losses can be found in Supplementary Table 9.

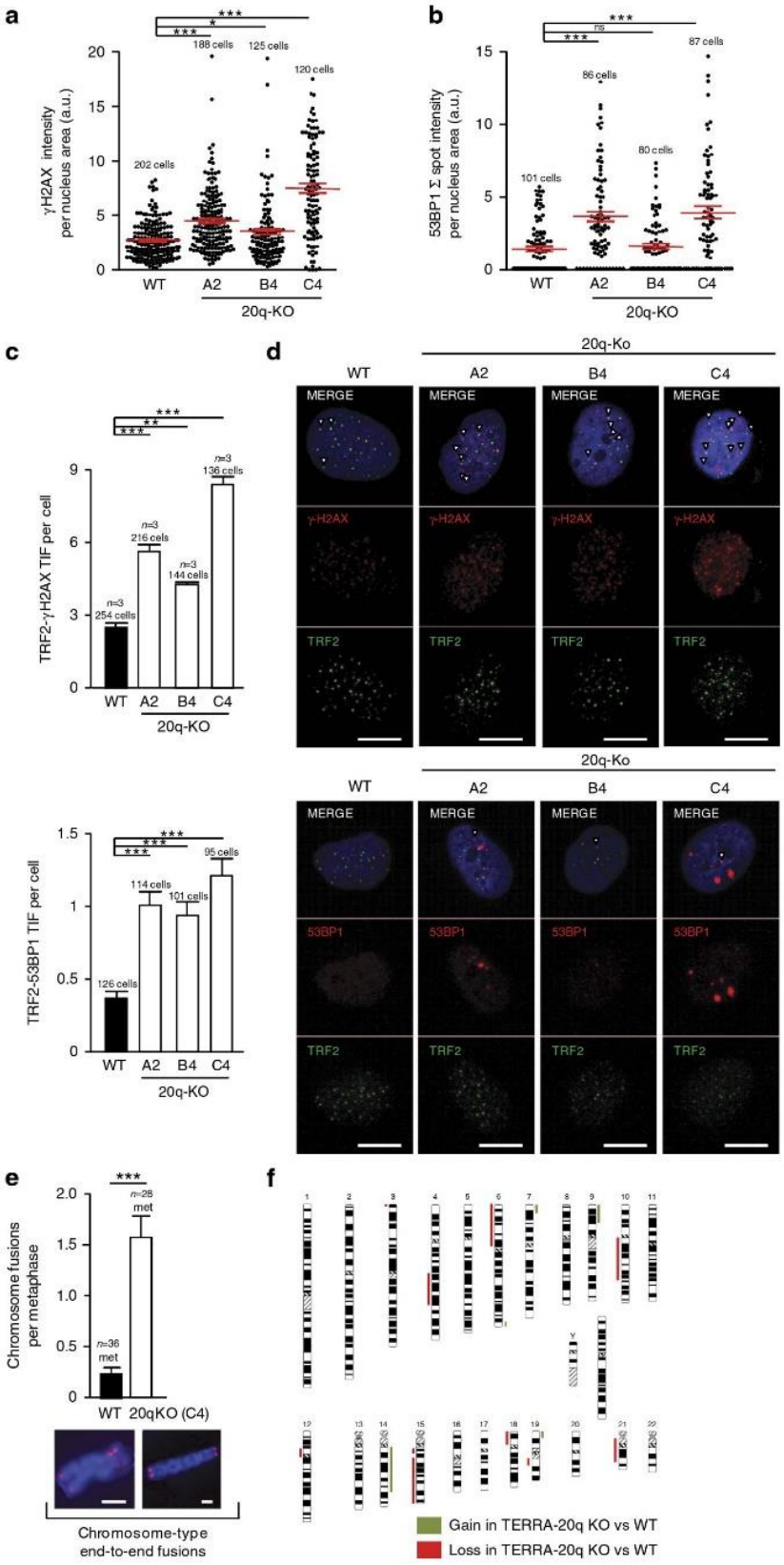
Finally, to ascertain whether TERRA deletion led to changes in the telomere binding proteins or shelterins, we determined the amounts of TRF1 and TRF2 telomere binding proteins at telomeres by either ChIP or immunofluorescence (Supplementary Fig. 6A,B). We observed a reduced abundance of both TRF1 and TRF2 at telomeres by ChIP in all 20q-KO clones compared with WT controls, which was more dramatic in the C4 clone, in agreement with shorter telomeres and higher chromosomal instability in this clone (Supplementary Fig. 6A). The 20q-KO C4 clone also showed a significant reduction in TRF1 and TRF2 fluorescence (Supplementary Fig. 6B), in agreement with increased telomere damage and telomere aberrations[23].

Transcript identity between 20q-TERRA- and Xp-RNAs. Given the fact that we observed co-localization of the transcripts arising from the Xp locus with TERRA, we next tested the possibility of sequence similarities between the 20q-TERRA transcripts and those arising from the Xp locus. To that end, we use the RNA-seq data from Porro *et al.*,[13] to generate modelled-transcripts (see 'Methods' section) for the 20q and Xp loci (Supplementary Fig. 7). We next performed pairwise alignments between the modelled-transcripts from each locus to calculate the percentage of identity between them (Supplementary Fig. 7; Supplementary Table 10). By doing this, we found 50–80% identity between some of the 20q and Xp modelled-transcripts (Supplementary Table 10). However, we cannot rule out that this transcript similarity is due to either the small transcript size or to the presence of repeats within the sequence of the modelled-transcripts (there are multiple annotated repeats within the transcribing regions of the modelled-transcripts showing higher identity) or to both (Supplementary Table 11). To counteract the transcript size effect, we bioinformatically generated one single-transcript for each of the 20q and Xp regions that encompasses all the previously modelled-transcripts for each region (see above and Supplementary Table 12). Next, we performed pairwise comparison between the single-transcripts corresponding to each region. As expected, the percentage of identity between the 20q and Xp single-transcripts decreased to $<40\%$, thus confirming the dependency of the sequence identity with the transcript size (Supplementary Fig. 7D).

We also found that transcripts arising from the Xp locus are more abundant than those arising from the 20q-TERRA locus (Supplementary Fig. 8A). Interestingly, the transcripts arising from the 20q locus were dramatically increased on deletion of the Xp locus as determined with three different pairs of primers designed within the 20q locus (Supplementary Fig. 8B). This

Figure 4 | Deletion of the 20q-TERRA locus decreases telomere length and protection in U2OS cells. (**a**) Q-FISH images obtained from metaphases spreads from U2OS cells WT and KO for the Chr20q-TERRA locus (clones A4, B4 and C4). (Left graphs) Frequency graphs of telomere length (a.u.) distribution measured in WT and in the 20q-KO cells (clones A4, B4 and C4) from three independent experiments. The mean telomere length and the number of telomeres and metaphases analyzed is shown. The red lines are arbitrary lines placed in the exact same position in each frequency graph to visualize differences between the 20q-KO clones and the WT controls (right graphs) The mean telomere length, the percentage of short telomeres and the quantification of signal-free ends per metaphase are also represented. Short telomeres are considered those in the 10% percentile of the total telomere length distribution. Total number of metaphases used for the statistical analysis is indicated. Scale bar, 10 μm and (zoom) 1 μm. (**b**) WT and 20q-KO cells were analyzed for T-SCE events with G-rich (green) and C-rich (red) PNA probes. The fraction of chromosome ends with T-SCE obtained from three different experiments was quantified and graphed as the mean values ± s.e.m., $n=30$ metaphases. The number of metaphases analyzed is shown. Only events in which interchange of both colours were quantified (see examples of no-T-SCE and T-SCE). The quantification was carried out by counting the number of events in the same chromosome or in different chromosomes and then normalizing it by the total number of chromosomes observed in each metaphase. Scale bar, 1 μm. (**c**) Quantification of DNA-containing double minute chromosomes (TDMs) in WT and 20q-KO cells from three different experiments (mean values ± s.e.m., $n=30$ metaphases). An example of TDMs is shown. One-way Anova with Dunnett's post test was used for all statistical analysis ($^*P<0.05$, $^{**}P<0.01$ and $^{***}P<0.001$). Scale bar, 1 μm.

finding is consistent with a compensation by the 20q-TERRA transcripts on deletion of the Xp locus explaining the absence of effects on total TERRA levels on deletion of the Xp locus (Fig. 3a,b). Deletion of the 20q locus also increased the Xp transcripts, mainly in the region in which the Xp deletion was generated (primer Xp/qPCR2) but not in the Xp region that was not deleted (primer Xp/qPCR1) (Supplementary Fig. 8B). Although these findings suggest a direct influence of the



transcription of the 20q-TERRA locus on the Xp locus and vice versa, with the current Xp CRISPR deletion in the clone D8 we cannot conclude that the transcripts arising from the Xp locus are genuine TERRA transcripts. Additional studies will be needed to understand whether the transcripts arising from the Xp contribute to TERRA.

Discussion

Here, we identified a major locus for human TERRA transcription at the subtelomere of the chromosome 20q. The genetic deletion of this locus allowed us to demonstrate for the first time in any organism the essential role of TERRA in the maintenance of functional telomeres.

In a previous report, eighteen different human subtelomeres were proposed to contain TERRA loci, although the authors did not study whether these transcripts had indeed TERRA features[13]. Here, we found that only two out of the eighteen loci have TERRA features, in particular, those corresponding to 20q and Xp subtelomeres. Furthermore, on genetic deletion of these loci by using CRISPR-Cas9, only the 20q locus was confirmed to be a bona fide TERRA locus since it resulted in almost complete abrogation of TERRA expression. Failure to decrease TERRA levels on Xp deletion may be explained by the impossibility to delete the entire Xp locus due to the highly repetitive nature of this region, or may indicate that this is not a bona fide TERRA locus. Alternatively, our finding that 20q-TERRAs are increased on deletion of the Xp locus, may also explain the fact that Xp deletion did not result in decreased TERRA levels. In any case, the fact that deletion of the 20q locus results in dramatically reduced TERRA levels demonstrates that this locus is a major TERRA locus in human cells.

Interestingly, the finding of one (or two at most) TERRA locus in human cells resembles the situation recently described by us in the case of mouse TERRA[3]. In particular, by using a similar approach to that used in humans (RNA-seq following UUAGGG-transcript enrichment), we identified mouse transcripts arising from thirteen different chromosome ends but only those transcribed from the subtelomere of chromosome 18 and 9 were confirmed to be bona fide TERRA transcripts, being the chromosome 18 locus responsible for the majority of TERRA transcripts[3]. Thus, both in human and mouse, TERRAs are transcribed from one or two subtelomeres, a finding that was not anticipated since it was widely accepted that TERRA were arising from all chromosome ends. It remains to be deciphered whether TERRA transcripts play different roles depending on the locus of origin.

Importantly, here we describe the first genetic deletion of TERRA performed in any organism. The findings are striking as they clearly demonstrate an essential role of TERRA in cellular viability as well as in telomere maintenance and telomere capping. In particular, the impossibility to expand most of the homozygous clones for the 20q-deletion in three independent human cell lines (Hela, HCT116 and U2OS) suggests an essential role of TERRA for cell survival in the majority of cell lines studied here. Only the U2OS cell line, which expresses higher basal levels of TERRA (this paper; refs 7,8), could bear the 20q-deletion suggesting that the cell needs to maintain a minimum amount of TERRA levels to survive. Importantly, we found a dramatic loss of telomeric sequences in U2OS cells on deletion of the 20q-TERRA locus, indicating that TERRA is essential for telomere maintenance. This dramatic telomere shortening was accompanied by telomere uncapping as indicated by increased telomeric damage and increased telomere fusions. Interestingly, U2OS cells are known to maintain telomeres by homologous recombination, a process in which TERRA has been suggested to participate[7–9]. In line with this, we found detectable but significantly decreased telomere recombination frequencies as indicated by T-SCE frequency. Together, these results indicate an essential role for TERRA in cellular viability and in the maintenance of telomeres, even in cells that maintain telomeres by recombination.

Interestingly, deletions and amplifications of the 20q subtelomere have been found in patients with different types of hematopoietic malignancies and with mental retardation[24–27]. Although some of these deletions also include coding genes, it is tempting to speculate a specific role of the 20q-TERRA on these malignancies. Finally, the finding of the human 20q-TERRA locus described here warrants future studies to analyze of a possible association of 20q-TERRA to these and other human diseases.

Methods

Cells and treatments and transfection. Human HCT116, HeLa, U2OS (ATCC) and HCT116 DKO[17] were cultured according to the ATCC's recommendations. 5′ Azacytidine treatment was performed at 1 mM for 3 days. Plasmids were transfected with Lipofectamine 2000 (Thermo Scientific) except for the IMR90 cell line that was electroporated using the Neon Transfection System (Invitrogen), in both cases following the manufacturers' protocol.

RNA-deep sequencing alignment. Raw data from RNA-seq samples was downloaded from GEO (GSE56727) and analyzed with the nextpresso pipeline, as follows: sequencing quality was checked with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Due to a decrease in nucleotide quality at the end of the reads across the samples, the last 31 nucleotides were trimmed. Reads were then aligned to the human genome (GRCh37/hg19) with TopHat-2.0.10 (ref. 28), using Bowtie 1.0.0 (ref. 29) and Samtools 0.1.19 (ref. 30), allowing 4 mismatches and 5 multihits. In a similar way, reads were aligned to the human subtelomeric reference sequences downloaded from the H.C. Riethman lab web page (http://www.wistar.org/lab/harold-c-riethman-phd/page/subtelomere-assemblies).

RNA-FISH. Cells grown on poly-L-lysine-coated coverslips (Becton Dickinson) were placed in cytobuffer (100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 10 mM pipes pH 6.8) for 30 s, washed in cytobuffer with 0.5% Triton X-100 for 30 s, washed in cytobuffer for 30 s and then fixed for 10 min in 4% paraformaldehyde in phosphate-buffered saline (PBS). The cells were dehydrated in 70, 80, 95 and 100% ethanol, air dried and hybridized overnight at 50 °C with RNA probes in

Figure 5 | Deletion of the 20q-TERRA locus decreases telomere protection in U2OS cells. (**a**) Quantification of the total γH2AX signal per nucleus (mean values ± s.e.m., n = number of cells) is shown. The total number of cells analyzed is indicated. (**b**) Quantification of the total 53BP1 spot signal per nucleus (mean values ± s.e.m., n = number of cells is shown). The total number of cells analyzed is indicated. (**c**) Graphs showing the quantification of the co-localization (TIF) between TRF2 and either γH2AX or 53BP1 in WT cells and in all 20q-KO clones (mean values ± s.e.m., $n = 3$ independent experiments for γH2AX and n = number of cells for 53BP1) per cell is shown. The total number of nuclei analyzed is indicated. (**d**) Representative images of the average number of TIFs found on double inmunostain to detect the telomere protein TRF2 (green) and either the DNA damage markers phospho-Histone γH2AX or 53BP1 (red) in the U2OS cells WT or deleted for the 20q locus. Arrowheads indicate co-localization events. Scale bar, 10 μm. (**e**) Quantification of chromosomal end-to-end fusions in WT and in the 20q-KO cells from three independent experiments (mean values ± s.e.m., n = metaphases). Examples of end-to-end fusions are shown as well. Scale bar, 1 μm. (**f**) Array-CGH analysis was performed on hybridization on the same membrane of DNA differentially labelled from WT and 20q-KO cells. The chromosomal gains and losses in 20q-KO cells normalized by WT cells are represented. The chromosomal gains are shown in green and in red the chromosomal losses. One-way Anova with Dunnett's post test was used for all statistical analysis except for the quantification of chromosomal fusions in which the Student's t-test was used (*$P<0.05$, **$P<0.01$ and ***$P<0.001$).

hybridization buffer (2 × sodium saline citrate (SSC)/50% formamide). Coverslips were washed two times for 15 min in hybridization buffer at 53 °C, two times for 10 min in 2 × SSC at 53 °C, 10 min in 1 × SSC at 53 °C, 5 min in 4 × SSC at room temperature, 5 min in 4 × SSC containing 0.1% Tween-20 and DAPI (Molecular Probes) at room temperature and 5 min in 4 × SSC at room temperature. Signals were visualized in a confocal ultraspectral microscope SP5-WLL (Leica). TERRA signal was quantified using the Definiens Developer XD.2 software RNA-FISH probes were generated from PCR products by *in vitro* transcription (Ambion) using Cy3-labelled CTP (Amersham); primers available in Supplementary Table 13. Quantification of the percentage of co-localization in double RNA-FISH was done by counting the average number of co-localization events detected in the individual confocal layers with respect to the total number of foci per nucleus arising from the RNA to be quantified. The percentage of co-localization was then represented (Fig. 1d).

Northern blot and dot-blot. Northern blot and dot-blot analyses were performed using standard protocols. TERRA probe was obtained from a 1.6-kb (TTAGGG)*n* cDNA insert excised from pNYH3 (kind gift from T. de Lange, Rockefeller University, NY, USA). Northern was normalized using 18S probes and quantified using ImageJ. Probes against subtelomere transcripts were prepared from PCR products from total cDNA (see primers used in Supplementary Table 13). The probes were labelled using the commercial Random Prime Labelling System Rediprime II (Amersham).

Plasmids and reporter assays. For construction of the reporter plasmid, PCR products were prepared with primers spanning the promoter regions of interest (Supplementary Table 14) and cloned into the plasmid pGL3-basic (Promega). Firefly and Renilla Luciferase activities were measured in cells harvested 2 days after transfection with the Dual Luciferase Reporter Assay System (Promega), following the manufacturer's instructions.

Generation of TERRA KO clones using the CRISPR-Cas9 system. The sequence for each gRNA (Supplementary Table 5) were obtained using a online CRISPR Design Tool[18] and those with the best score for ruling out off-target mutations were chosen. Each gRNA was cloned into the plasmid pSpCas9(BB)-2A-GFP (PX458) (Gift from Feng Zhang) (Addgene plasmid #48138) using the protocol described in Hsu *et al.*,[18]. Cells were transiently co-transfected with two of this plasmid containing each one the Cas9 and a specific gRNA. Single cell clones were obtaining by sorting. To this end, two days after transfection cell were trypsinized and washed with dPBS, 1, 3, 5 or 10 cells of the 10% GFP brightest ones were sorted using the FACS ARIA IIU (Becton Dickinson) into 96-wells plates. Two weeks later, wells were checked for monoclonal cell expansion and that cells were genotyped by PCR. Homozygous monoclonal cell lines were expanded. PCR primer can be found in Supplementary Table 7.

PCR and Sanger sequencing. For sequencing, samples were amplified using the GoTaq polymerase (Promega) and cloned into the commercial plasmid pCR2.1-TOPO (Invitrogen) following the manufacturer's instructions. Samples were sequences using the Sanger-style BigDye terminator chemistry on an ABI 3730 ×l sequencer (Applied Biosystems).

Telomere QFISH and chromosomal aberrations on metaphases. Colcemide (Gibco) was added to cells at a concentration of 0.1 μg ml^{-1} during 4 h. After hypotonic swelling in 0.03 M sodium citrate for 25 min at 37 °C, cells were fixed in methanol:acetic acid (3:1). After 2–3 fixative changes, the cell suspension was dropped onto clean, wet microscope slides and dried overnight. Then, slides were washed with PBS and fixed for 2 min in 4% formaldehyde. Slides were washed three more times with PBS and treated with pepsin (1 mg ml^{-1}, pH 2) at 37 °C for 10 min. Formaldehyde fixation and washing steps were repeated, then slides were incubated in 0.4 M HCl for 10 min, rinsed in PBS, dehydrated in an ethanol series (70, 90 and 100%) and air dried. The Cy-3-labelled (CCCTAA)3 PNA probe (Panagene) was dissolved in hybridization buffer containing 70% formamide/10 mM Tris pH 7 and 0.25% (w/v) blocking reagent (0.5 μg ml^{-1}) (Roche). The hybridization mixture was placed onto the slides and a coverslip was applied, followed by DNA denaturation (3 min, 80 °C). After hybridization (2 h, room temperature), slides were washed twice with 70% formamide/10 mM Tris pH 7.2 for 15 min, followed by a 5 min washing in 0.05 M Tris/0.15 M NaCl pH 7.5/0.05% Tween-20. Slides were dehydrated in an ethanol series and air dried. Finally, slides were counterstained with Vectashield containing DAPI (Vector Laboratories, Burlingame, CA). The cut-off used in the determination of the percentage of short telomeres was determined by averaging the results obtained after subtracting the 10% percentile from the mean telomere length value in the WT condition. For analysis of chromosomal aberrations, metaphases were analyzed by superimposing the telomere image on the DAPI image using TFL-telo.

Immunofluorescence. Cells grown on poly-L-lysine-coated coverslips (Becton Dickinson) were placed in cytobuffer (100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 10 mM pipes pH 6.8) for 30 s, washed in cytobuffer with 0.5% Triton X-100 for 30 s, washed in cytobuffer for 30 s for and then fixed for 10 min in 4% paraformaldehyde in PBS. Cells were blocked with 100% FBS for 1 h at room temperature. Cells were incubated with primary antibody dissolved in Dako antibody Diluent (Dako) overnight in a humid chamber at 4 °C. Next day, coverslips were washed 3 times for 30 min in 1 × PBS containing 0.1% Tween-20. Cells were incubated with Alexa secondary antibody (Life Technologies, A11017) dissolved in Dako antibody Diluent (Dako) for 1 h in a humid chamber at room temperature. Cells were washed three times for 30 min in 1 × PBS. Samples were mounted in Prolong with Dapi (Invitrogen). Signals were visualized in a confocal ultraespectral microscope SP5-WLL (Leica). The following antibodies were used: anti-phospho-Histone γH2AX (05-636, Millipore) diluted 1:200, anti-53BP1 (NB-100-304, Novus) diluted 1:200, anti-TRF1 (TRF78, Abcam) diluted 1:200 and anti-TRF2 (clone 4A794, Millipore) diluted 1:200.

CGH array. Oligo array-CGH analysis was performed using an Agilent SurePrint G3 Human CGH 60 K microarray (AMADID 021924 Agilent Technologies, Santa Clara, CA, USA) spanning the entire human genome at a median resolution of 41 kb. Totally, 500 ng of genomic DNA from the 20q-KO and WT cells were differentially labelled by random priming with Cy5-dCTP and Cy3-dCTP. The hybridization was carried out according to the manufacturer's protocol. Arrays were scanned using an Agilent DNA Microarray scanner G2565CA (Agilent Technologies). Microarray data were extracted and visualized using Feature Extraction software v10.7 and Agilent Genomic Workbench (AGW) software v7.0 (Agilent Technologies). Copy number altered regions were detected using the Aberration Detection Method 2 (ADM-2) algorithm set as 6, with a minimum number of three consecutive probes.

Chromosome orientation FISH (CO-FISH). Exponentially growing cells were sub-cultured in the presence of 5′-bromo-2′-deoxyuridine (BrdU; Sigma) at a final concentration of 1×10^{-5} M, and then allowed to replicate their DNA once at 37 °C for 24 h. Colcemide (Gibco) was added at a concentration of 0.1 μg ml^{-1} during the last 4 h. Cells were then recovered and metaphases prepared (see Q-FISH methods). The slides were treated with 0.5 mg ml^{-1} RNase A for 10 min at 37 °C, stained with 0.5 μg ml^{-1} Hoechst 33258 (Sigma) in 2 × SSC (0.3 M NaCl, 0.03 M sodium citrate) for 15 min at room temperature and then exposed to 365 nm UV light (Stratalinker 1800 UV irradiator) for 25–30 min. Enzymatic digestion of the BrdU/BrdC-substituted DNA strands with 3 U μl^{-1} of Exonuclease III (Promega) in buffer supplied by the manufacturer (50 mM Tris–HCl, 5 mM $MgCl_2$ and 5 mM dithiothreitol, pH 8) was allowed to proceed for 10 min at room temperature. An additional denaturation in 70% formamide, 2 × SSC at 70 °C for 1 min was performed in order to ensure complete removal of the newly replicated bromo-substituted strands and followed by dehydration in a cold ethanol series (70, 85, 100%). A hybridization mixture containing 0.4 μg ml^{-1} of CY3-labelled PNA probe against the telomeric lagging strand (Panagene) in 30% formamide and 2 × SSC was applied to slides. Following 4 h of hybridization in a moist chamber at 37 °C, the slides were washed 5 times for 15 min each in 2 × SSC at 42 °C. The same hybridization step was performed with a hybridization mixture containing 0.4 μg ml^{-1} of Fluoresceine-labelled PNA probe against the telomeric leading strand (Panagene). Finally, slides were counterstained with Vectashield containing DAPI (Vector Laboratories, Burlingame, CA, USA). T-SCE events were counted when interchange of both probes was seen.

Bioinformatic analysis of the probe specificity. The bioinformatic prediction of the regions amplified by the different primers to generate the probes was calculated using the *in silico* PCR tool from the UCSC human genome browser GRCh38/hg38. The homology of the best amplicon for each pair of primers was calculated using the NCBI nucleotide BLAST aligning the sequence of the first 15-kb from each subtelomere sequence obtained from the subtelomeric browser from Stong *et al.*,[14] against all the possible amplicons sequence.

Transcript identity studies. Reads from RNA-seq samples[13] were aligned to the human subtelomeric reference sequences downloaded from the H.C. Riethman lab web page (http://www.wistar.org/lab/harold-c-riethman-phd/page/subtelomere-assemblies). To study transcript identity, alignments were used to obtain transcripts modelled by Cufflinks inside the chr.20q and chr.Xp subtelomeric regions of interest (corresponding to the regions to be deleted with CRISPR/Cas9) and their FASTA sequences retrieved. A one-exon transcript was generated that spans all the different transcripts modelled by Cufflinks for each of the conditions. Next, we performed pairwise alignments among them, using either the Emboss Emboss Water tool or the Emboss Needle tool[32], to calculate the local or global sequence identity, respectively. Local comparison was performed for the modelled-transcripts and global comparison for the one-exon transcripts.

Off-target mutation analysis. DNA regions bearing predicted off-target mutations were amplified by PCR (see primers in Supplementary Table 15) and TIDE analysis[32]. DNA from cells that were not transfected with CRISPR-Cas9 was used as reference. T7-endonuclease assay was performed as follows: a DNA

heteroduplex formation was perfomed in a 20 µl reaction containing 300 ng of DNA and 2 µl of buffer 2 (NEB). The conditions for the reaction of the DNA heteroduplex formation are indicated in Ran *et al.*[32]. On DNA heteroduplex formation, digestion was performed with 2 µl of endonuclease T7 (NEB) for 30 min at 37 °C. The digestion was stopped with EDTA and the DNA products visualized in 2% agarose (Nusieve) gels.

Raw gels and blots. Full gels and blots from the main figures can be found in Supplementary Figure 9.

Data availability. Raw data from RNA-seq samples was downloaded from GEO (GSE56727). All relevant data are available from the authors.

References

1. Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318,** 798–801 (2007).
2. Schoeftner, S. & Blasco, M. A. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat. Cell Biol.* **10,** 228–236 (2008).
3. Lopez de Silanes, I. *et al.* Identification of TERRA locus unveils a telomere protection role through association to nearly all chromosomes. *Nat. Commun.* **5,** 4723 (2014).
4. Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. & Lieberman, P. M. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol. Cell* **35,** 403–413 (2009).
5. Lopez de Silanes, I., Stagno d'Alcontres, M. & Blasco, M. A. TERRA transcripts are bound by a complex array of RNA-binding proteins. *Nat. Commun.* **1,** 33 (2010).
6. Flynn, R. L. *et al.* TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature* **471,** 532–536 (2011).
7. Flynn, R. L. *et al.* Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. *Science* **347,** 273–277 (2015).
8. Arora, R. *et al.* RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat. Commun.* **5,** 5220 (2014).
9. Balk, B. *et al.* Telomeric RNA–DNA hybrids affect telomere-length dynamics and senescence. *Nat. Struct. Mol. Biol.* **20,** 1199–1205 (2013).
10. Redon, S., Reichenbach, P. & Lingner, J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res.* **38,** 5797–5806 (2010).
11. Cusanelli, E., Romero, C. A. & Chartrand, P. Telomeric noncoding RNA TERRA is induced by telomere shortening to nucleate telomerase molecules at short telomeres. *Mol. Cell* **51,** 780–791 (2013).
12. Nergadze, S. G. *et al.* CpG-island promoters drive transcription of human telomeres. *RNA* **15,** 2186–2194 (2009).
13. Porro, A. *et al.* Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat. Commun.* **5,** 5379 (2014).
14. Stong, N. *et al.* Subtelomeric CTCF and cohesin binding site organization using improved subtelomere assemblies and a novel annotation pipeline. *Genome Res.* **24,** 1039–1050 (2014).
15. Costa, V. *et al.* DDX11L: a novel transcript family emerging from human subtelomeric regions. *BMC Genomics* **10,** 250 (2009).
16. Linardopoulou, E. V. *et al.* Human subtelomeric WASH genes encode a new subclass of the WASP family. *PLoS Genet.* **3,** e237 (2007).
17. Rhee, I. *et al.* DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* **416,** 552–556 (2002).
18. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–832 (2013).
19. Conomos, D., Pickett, H. A. & Reddel, R. R. Alternative lengthening of telomeres: remodeling the telomere architecture. *Front. Oncol.* **3,** 27 (2013).
20. Draskovic, I. & Londono Vallejo, A. Telomere recombination and alternative telomere lengthening mechanisms. *Front. Biosci. (Landmark Ed.)* **18,** 1–20 (2013).
21. Bailey, S. M., Brenneman, M. A. & Goodwin, E. H. Frequent recombination in telomeric DNA may extend the proliferative life of telomerase-negative cells. *Nucleic Acids Res.* **32,** 3743–3751 (2004).
22. Zhu, X. D. *et al.* ERCC1/XPF removes the 3′ overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. *Mol. Cell* **12,** 1489–1498 (2003).
23. Bandaria, J. N., Qin, P., Berk, V., Chu, S. & Yildiz, A. Shelterin protects chromosome ends by compacting telomeric chromatin. *Cell* **164,** 735–746 (2016).
24. Courtens, W. *et al.* A de novo subtelomeric monosomy 11q (11q24.2-qter) and trisomy 20q (20q13.3-qter) in a girl with findings compatible with Jacobsen syndrome: case report and review. *Clin. Dysmorphol.* **16,** 231–239 (2007).
25. Gray, B. A., Cornfield, D., Bent-Williams, A. & Zori, R. T. Translocation (X;20)(q13.1;q13.3) as a primary chromosomal finding in two patients with myelocytic disorders. *Cancer Genet. Cytogenet.* **141,** 169–174 (2003).
26. Kroepfl, T., Petek, E., Schwarzbraun, T., Kroisel, P. M. & Plecko, B. Mental retardation in a girl with a subtelomeric deletion on chromosome 20q and complete deletion of the myelin transcription factor 1 gene (MYT1). *Clin. Genet.* **73,** 492–495 (2008).
27. Okada, M. *et al.* Microarray CGH analyses of chromosomal 20q deletions in patients with hematopoietic malignancies. *Cancer Genet.* **205,** 18–24 (2012).
28. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7,** 562–578 (2012).
29. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
30. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
31. Li, W. *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **43,** W580–W584 (2015).
32. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–2308 (2013).
33. Llanos, S., Efeyan, A., Monsech, J., Dominguez, O. & Serrano, M. A high-throughput loss-of-function screening identifies novel p53 regulators. *Cell Cycle* **5,** 1880–1885 (2006).

Acknowledgements

We are indebted to R. Torres and S. Rodriguez for advice in the CRISPR-Cas9 technology and helpful discussions. We thank D. Megías for the quantification of confocal image signals and to O. Dominguez for troubleshooting with cloning and sequencing. We thank S. Llanos for the p21 promoter reporter (described in ref. 33). Research in the Blasco lab is funded by the Spanish Ministry of Economy and Competitiveness Project (SAF2013-45111-R), the Madrid Regional Government Project S2010/BMD-2303 (ReCaRe), Fundación Botín (Spain) and AXA Research Fund (AXA 2011, Spain).

Author contributions

M.A.B. conceived the original idea. J.J.M. and I.L.d.S. conducted and analyzed the experiments and O.G. did the bioinformatic analysis. M.A.B., J.J.M and I.L.d.S. designed the experiments. M.A.B and I.L.d.S. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interest.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Montero, J.J. *et al.* Telomeric RNAs are essential to maintain telomeres. *Nat. Commun.* 7:12534 doi: 10.1038/ncomms12534 (2016).

Artículo 2

TERRA recruitment of polycomb to telomeres is essential for histone trymethylation marks at telomeric heterochromatin

Artículo 2: TERRA recruitment of polycomb to telomeres is essential for histone trymethylation marks at telomeric heterochromatin.

Autores: Juan José Montero, Isabel López-Silanes, Diego Megías, Mario F. Fraga, Alvaro Castells-García y María A. Blasco

Publicado en Nature Communications. 18 de abril de 2018. 9, número de artículo: 1548

Resumen:

Estudios previos habían puesto de manifiesto una posible implicación de TERRA en la regulación de la cromatina telomérica. Estos estudios indicaban que TERRA sería importante para la compactación del telómero a través de la interacción con diferentes proteínas como HP1. Para estudiar el papel de TERRA en el establecimiento de la cromatina telomérica, usamos celulares de U2OS 20q-TERRA KO. Para ello generamos nuevos clones sin la región del subtelómero 20q que contiene el origen transcripcional de TERRA células

Posteriormente evaluamos el efecto de TERRA en el establecimiento de la heterocromatina telomérica. Observamos que TERRA e importante para la deposición de las marcas de heterocromatina asociadas al telómero (H3K9me3, H4K20me3 y HP1), mientras que no es necesario para el depósito de marcas de eucromatina. Además, encontramos que TERRA es importante para la deposición de la marca H3K27me3 en el telómero, una marca que es sintetizada exclusivamente por el complejo PRC2 (*Polycomb repressive complex 2*)

Posteriormente, identificamos que TERRA interacciona con el complejo PRC2 y que PRC2 es capaz de unirse al telómero. Comprobamos que esta unión es dependiente de TERRA, sugiriendo que el papel de TERRA en el establecimiento de la cromatina telomérica podría estar mediado por el complejo PRC2. Para validar esta hipótesis, generamos *Knock-down* de distintas subunidades del complejo PRC2 en las células U2OS. Vimos que la eliminación de las subunidades del complejo PRC2 causaba la pérdida de las marcas de heterocromatina (H3K9me3, H4K20me3, H3K27me3 y HP1) en el telómero.

En conclusión, este estudio demuestra un papel de TERRA en el mantenimiento de la heterocromatina telomérica y que esta función esta mediada por el complejo PRC2.

Artículo 2

Contribución Personal: He participado en el diseño del estudio, realizado y analizado la mayoría de los experimentos (clonaje del sistema CRISPR y del shRNAs, generación y confirmación de las líneas celulares TERRA KO y silenciamientos, análisis de la longitud telomérica, ChIP dot-blot telomérico, y evaluación de la interacción de PRC2 con TERRA y el telómero). Finalmente, he contribuido en la discusión de los resultados y en la preparación del artículo al elaborar parte del texto y las figuras.

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-03916-3 OPEN

TERRA recruitment of polycomb to telomeres is essential for histone trymethylation marks at telomeric heterochromatin

Juan J. Montero[1], Isabel López-Silanes[1], Diego Megías[2], Mario F. Fraga[3], Álvaro Castells-García[4,5] & Maria A. Blasco[1]

TERRAs are long non-coding RNAs generated from the telomeres. Lack of TERRA knockout models has hampered understanding TERRAs' functions. We recently identified chromosome 20q as one of the main origins of human TERRAs, allowing us to generate the first 20q-TERRA knockout models and to demonstrate that TERRAs are essential for telomere length maintenance and protection. Here, we use ALT 20q-TERRA knockout cells to address a direct role of TERRAs in telomeric heterochromatin formation. We find that 20q-TERRAs are essential for the establishment of H3K9me3, H4K20me3, and H3K27me3 heterochromatin marks at telomeres. At the mechanistic level, we find that TERRAs bind to PRC2, responsible for catalyzing H3K27 tri-methylation, and that its localization to telomeres is TERRA-dependent. We further demonstrate that PRC2-dependent H3K27me3 at telomeres is required for the establishment of H3K9me3, H4K20me3, and HP1 binding at telomeres. Together, these findings demonstrate an important role for TERRAs in telomeric heterochromatin assembly.

[1] Telomeres and Telomerase Group, Molecular Oncology Program, Melchor Fernández Almagro 3, E-28029 Madrid, Spain. [2] Confocal Microscopy Unit, Spanish National Cancer Centre (CNIO), Melchor Fernández Almagro 3, E-28029 Madrid, Spain. [3] Cancer Epigenetics Laboratory, Nanomaterials and Nanotechnology Research Center (CINN-CSIC)-Universidad de Oviedo, Institute of Oncology of Asturias (IUOPA) and Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Avda De la Vega, 4-6, 33940 El Entrego, Spain. [4] Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Drive Aiguader 88, 08003 Barcelona, Spain. [5] Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain. Correspondence and requests for materials should be addressed to M.A.B. (email: mblasco@cnio.es)

Telomeres are nucleoprotein structures at the ends of chromosomes that protect them from being recognized as DNA double-strand breaks, thus preventing chromosomal end-to-end fusions[1]. In vertebrates, telomeres consist of tandem repeats of the TTAGGG sequence bound by the so-called telomere-binding proteins or shelterin, which are essential for the formation of a functional telomere cap[2]. Telomere repeats can be generated de novo by telomerase, a reverse transcriptase (*TERT*) that elongates chromosome ends by using an RNA component (*TERC*) as template[1]. In mice and humans, telomerase is highly expressed in embryonic pluripotent stem cells, and this expression is downregulated after birth leading to progressive telomere erosion with aging owing to the incomplete replication of linear chromosomes[3,4]. In contrast, cancer cells aberrantly overexpress telomerase allowing for the ability of cancer cells to proliferate indefinitely[5]. Cancer cells can also use an independent mechanism to elongate telomeres known as alternative lengthening of telomeres (ALT), that is based on homologous recombination between telomeric sequences[6].

Interestingly, both telomere length and telomere recombination are subjected to a higher-order regulation involving epigenetic modifications of the telomeric chromatin[7]. In particular, mammalian telomeres are enriched in heterochromatic marks, including HP1 binding, H3K9 and H4K20 tri-methylation histone marks, as well as hypermethylation of subtelomeric DNA[7–13]. These marks have been proposed to negatively regulate telomere length and telomere recombination[7]. In particular, cells lacking the histone methyltransferases Suv39 or Suv420, as well as cells lacking the DNA methyltransferases DNMT1 and 3 have markedly elongated telomeres[9,11,12]. Thus, disruption of this silent chromatin environment results in loss of telomere-length control and in increased telomere recombination. In turn, our group also showed that progressive telomere loss associated to cell division reduces chromatin compactation at telomeric and subtelomeric domains, which may favor telomere-elongation mechanisms[7,8,13]. Telomere chromatin also shows a decrease of heterochromatic histone marks during nuclear reprogramming[14], coinciding with net telomere elongation by telomerase[14].

Despite the heterochromatic environment of the telomere, telomeres are transcribed giving rise to long non-coding UUAGGG-repeat transcripts known as Telomeric repeat-containing RNAs (TERRAs) or TelRNAs[15,16]. TERRAs are transcribed from the subtelomere toward the telomere and show a spotted nuclear pattern as detected by RNA-FISH[15,16]. A proportion of TERRAs spots colocalizes with telomeres, suggesting that TERRAs are part of the telomeric chromatin[15,16]. At telomeres, TERRAs are essential to maintain telomere length and telomere protection[17–20], and therefore can be considered as bone fide telomere components. TERRAs have been also recently found to bind to extratelomeric regions, explaining the non-telomeric TERRA spots[17].

TERRAs have been implicated in telomere protection[17–21], heterochromatin formation[21], telomere replication[22,23], and in telomere elongation by homologous recombination through the formation of DNA-TERRA hybrids[24–26]. However, understanding of TERRA in vivo function and the direct involvement of TERRA in all the above-mentioned processes is largely pending owing to lack of loss-of-function TERRA models. In this regard, we recently identified the mouse and the human TERRA locus[18,20]. We found that in both human and mouse, TERRAs do not arise from all chromosomes but from a few of them[18,20]. In human cells, the majority of TERRAs arise from a single locus in chromosome 20q[20]. This allowed us to first generate knockout cells for this locus, which showed markedly reduced TERRA levels, thus demonstrating that 20q is the main origin of human TERRAs. 20q-TERRA knockout cells also demonstrated that TERRAs are essential for telomere maintenance and telomere protection[20]. Here, we use this human TERRA loss of function model to assess a direct role of TERRA long non-coding RNAs in telomeric heterochromatin formation, as well as unveil the underlying molecular mechanisms.

In particular, there is mounting evidence that non-coding RNAs are involved in heterochromatin formation across species both through *cis* and *trans* mechanisms. This includes dosage compensation in mammals, imprinting, and demarcation of gene-silencing chromosomal domains[27,28]. In addition, non-coding RNAs transcribed at centromeres are also proposed to be involved in higher-order chromatin structures, and their transcription to be important for the deposition of CenH3 (homologous centromere-specific histone H3 variants; ref. [29]).

In the case of TERRAs, not only do they colocalize with telomeric chromatin in *cis* and *trans*[15,16,18], but they can also form DNA–RNA hybrids or R-loops at the telomere[24–26], a type of structure shown to epigenetically modify the genome[30,31]. TERRAs have also been shown to interact with HP1 and H3K9me3[21]. Intriguingly, TERRAs have been proposed to correlate with H3K27 deposition at extratelomeric sites, a histone mark placed by EZH2, the catalytic subunit of the Polycomb repressive complex 2 (PRC2). However, whether TERRAs regulate H3K27 deposition at telomeres or, more generally, whether TERRAs have a direct role in the establishment of telomeric chromatin status is largely unknown owing to the lack of KO models for TERRAs.

Here, we set to address a direct involvement of TERRAs in the regulation of a higher-order regulation of chromatin at telomeres by using a panel of human 20q-TERRA KO cells with markedly reduced TERRA levels. We find that depletion of TERRAs in human U2OS cells cause a strong loss of H3K9me3 and H4K20me3 tri-methylation histone marks at telomeric chromatin, demonstrating direct involvement of TERRAs in the establishment of telomeric heterochromatin. In addition, we make the finding that human telomeres are enriched for the H3K27m3 trimethylation facultative heterochromatin mark, which is catalyzed by the PRC2 complex, and that this mark is also decreased in 20q-TERRA KO cells. Indeed, we describe here that TERRAs directly bind the PRC2 complex components EZH2 and SUZ12 and that this binding is critical for PRC2 recruitment to telomeres. We further demonstrate that PRC2 at telomeres is required for the establishment of H3K27me3, H3K9me3, H4K20me3, and HP1 binding at telomeres. In summary, we describe for the first time a role for PRC2 in the establishment of telomeric chromatin, which is regulated by TERRAs, thus demonstrating an important role for TERRAs in telomeric heterochromatin assembly.

Results

Generation of a panel of 20q-TERRA KO cells. We first generated a large panel of human U2OS cells KO for the 20q-TERRA locus with the CRISPR-Cas9 system using the same strategy described by us[20]. To increase the efficacy of deletion we used an all-in-one plasmid system[32], which contains the two gRNAs (E1 and S2) under the U6 promoter, the Cas9 protein, and a GFP cassette (see Materials and Methods; Fig.1a). EGFP-positive cells were sorted by flow cytometry and after PCR screening, we obtained eight clones that carried the deletion in homozygosis (clones #A7, E1, E6, E8, and H8; Fig. 1b). As controls, U2OS cells were electroporated with the same plasmid but without gRNAs (control clones #1 and #2). As expected, deletion of the 20q-TERRA locus resulted in a significant reduction in total TERRA levels in the deleted clones (A7, E1, E6, E8, and H8) compared with the original U2OS cellular pool and with the WT clones (#1 and 2; Fig. 1c), as determined by dot-blot analysis. Note that,

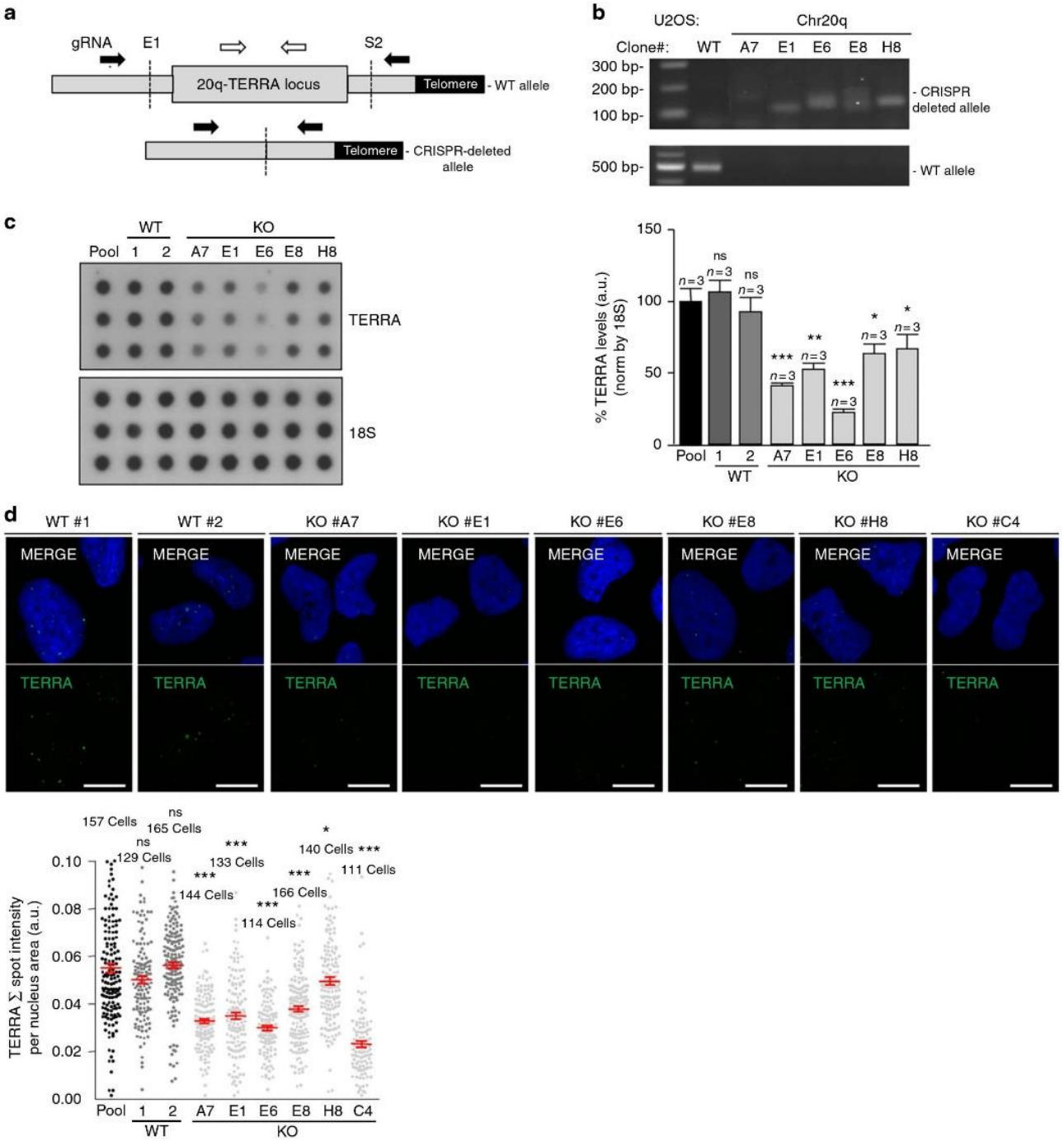


Fig. 1 Deletion of the TERRA-20q locus markedly affects TERRA expression. **a** Scheme depicting the WT and the CRISPR-deleted allele for the 20q-TERRA locus located in the subtelomere of the chromosome 20, q-arm. The position of the gRNAs (E1 and S2) and the primers used to genotype the deletions are also shown. The black arrows represent the primers to amplify the CRISPR-deleted allele and the white arrows the ones to amplify the WT allele located inside the 20q-TERRA locus. **b** Ethidium bromide gels showing the WT and the CRISPR-deleted allele for the 20q-TERRA locus detected by PCR in a WT cellular pool and in different clones of the U2OS cells. **c** RNA from the a WT cellular pool, WT expanded clones (#1 and 2), or 20q-TERRA KO clones (#A7, E1, E6, E8, and H8) from the USOS cell line was isolated and used for TERRA detection by RNA dot-blot with a probe against the TERRA-UUAGGG-tract; 18S serves as loading control. (Graph) TERRA quantification normalized by 18S (mean values ± s.e.m., $n = 3$ biological replicates). **d** Representative confocal microscopy images of RNA-FISH against TERRA-UUAGGG-tract (green) in the U2OS WT clones (#1 and 2) and in the 20q-TERRA KO clones (#A7, E1, E6, E8, H8, and C4). Scale bar, 10 μm. (Graph) Quantification of the total spot intensity per nucleus normalized by nucleus area (mean values ± s.e.m., n = cells analyzed). One-way ANOVA with Dunnett's post test was used for the statistical analysis (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$)

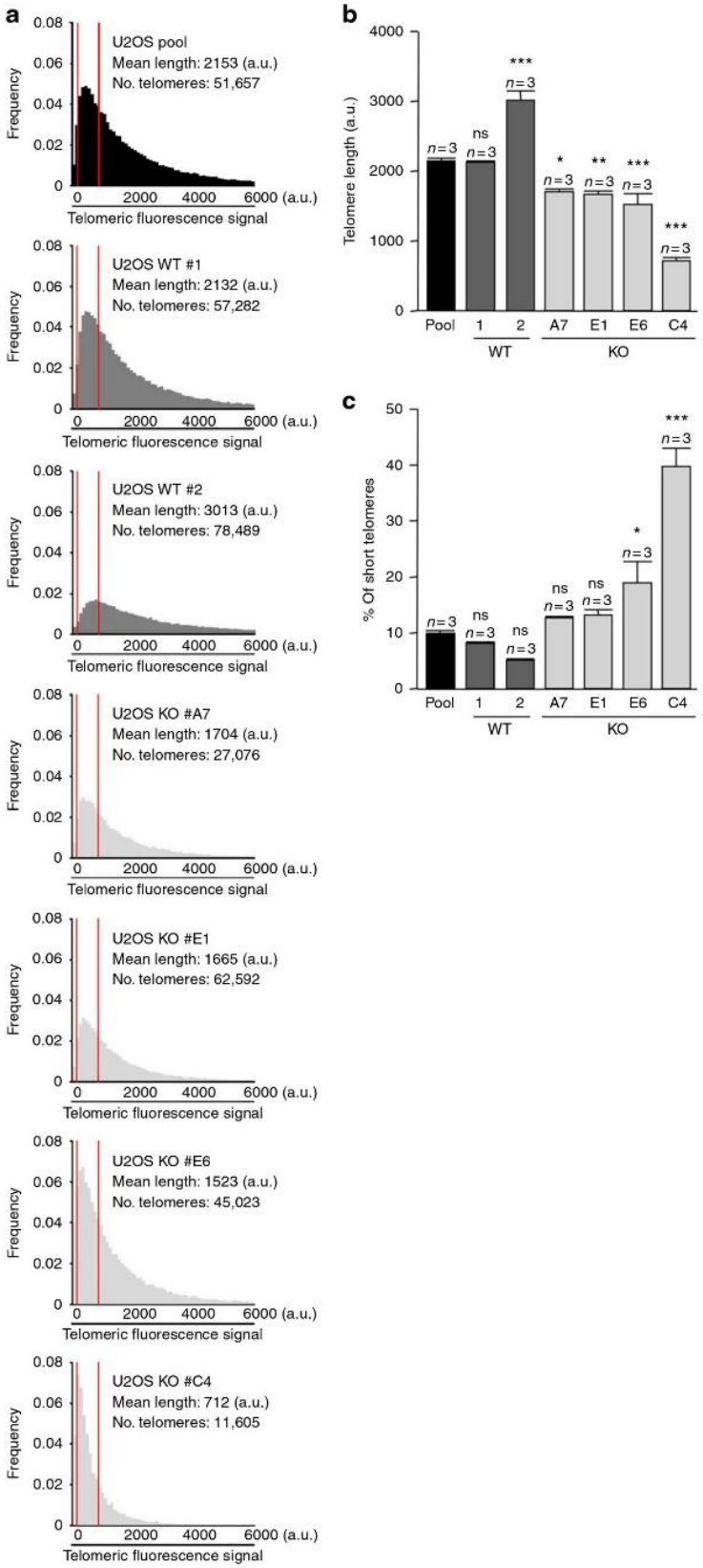


Fig. 2 Deletion of the 20q-TERRA locus decreases telomere length. **a** Representative frequency graphs of telomere length distribution (a.u.) measured in the U2OS WT pool (black), in the WT clones (#1 and 2; dark gray), and in the 20q-TERRA KO clones (#A7, E1, E6, and C4; light gray). The mean telomere length and the number of telomeres analyzed is shown. The red lines are arbitrary lines placed in the exact same position in each frequency graph to visualize differences between samples. **b** Graph showing the quantification of the mean telomere length in the U2OS cells WT pool, the WT expanded clones (#1 and 2), and in the 20q-TERRA KO clones (#A7, E1, E6, and C4) by HT-Q-FISH (mean values ± s.e.m., n = technical replicates). **c** Graph showing the percentage of short telomeres in the same settings. Short telomeres are considered those in the 10% percentile of the total telomere length distribution (mean values ± s.e.m., n = technical replicates). One-way ANOVA with Dunnett's post test was used for the statistical analysis (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$)

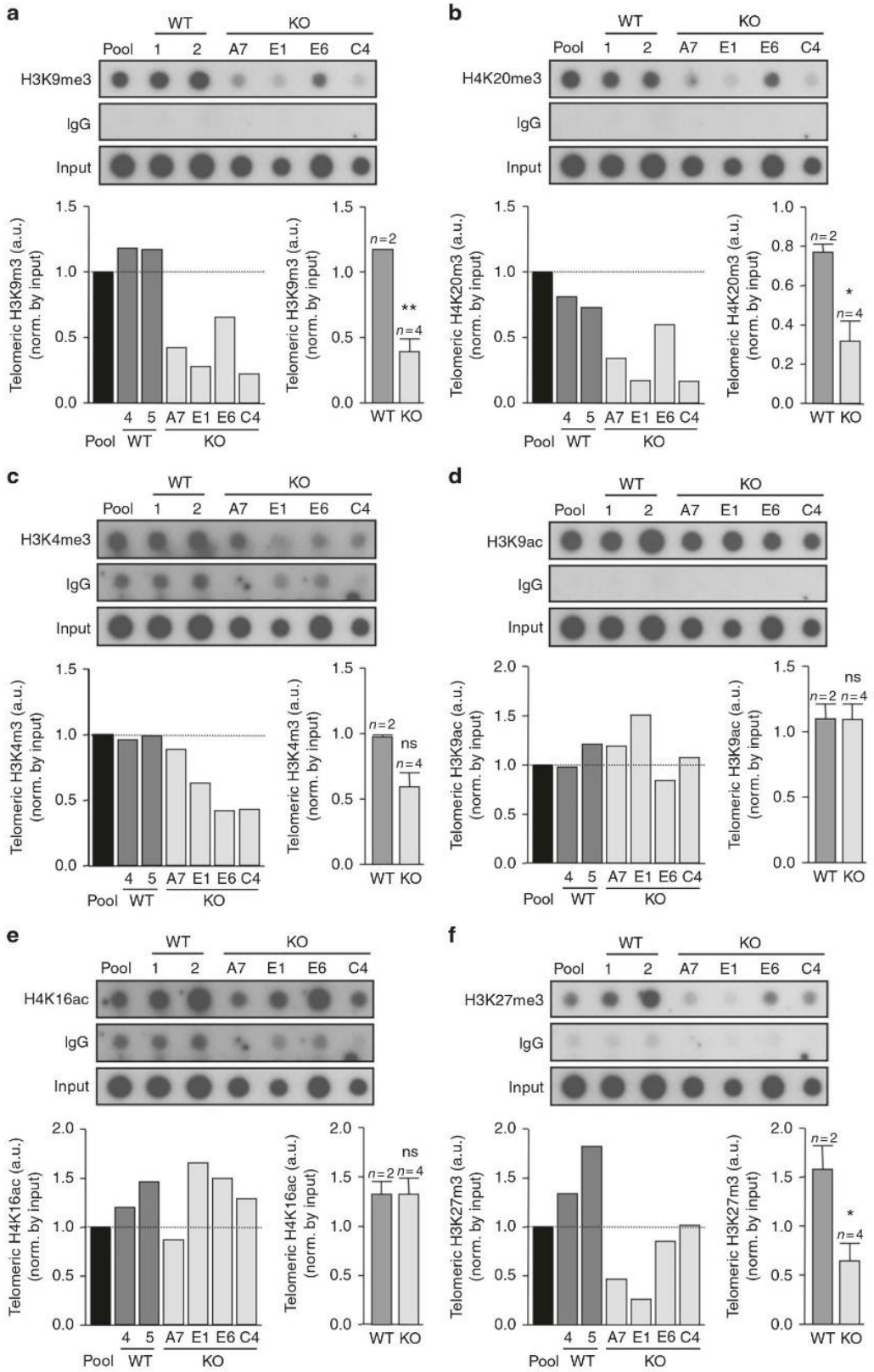
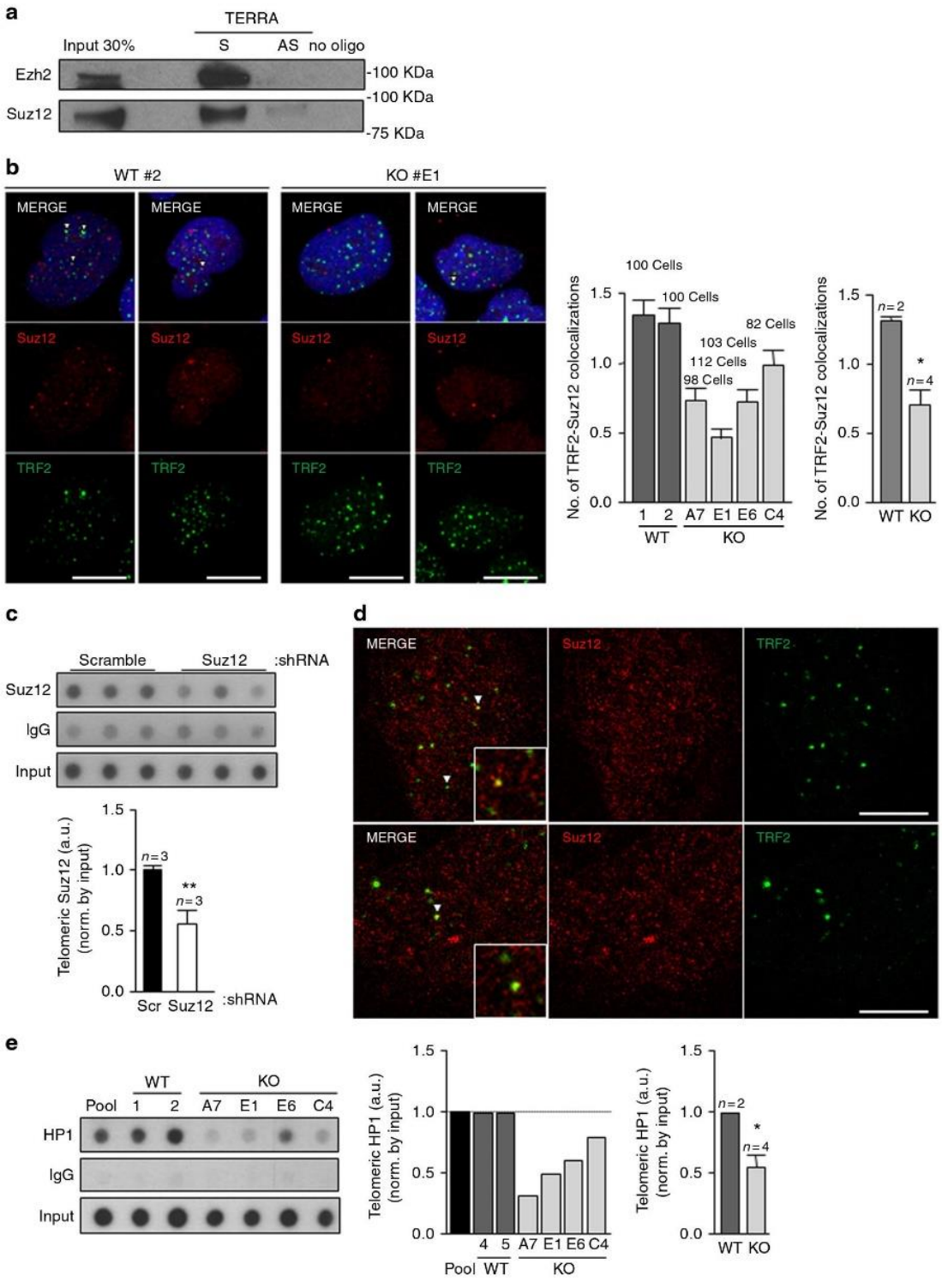


Fig. 3 TERRAs are essential for the assembly of heterochromatic histone marks at the telomere, including H3K27me3. **a** ChIP-dot-blot of the H3K9m3, **b** H4K20m3, **c** H3K4m3, **d** H3K9ac, **e** H4K16ac, and **f** H3K27me3 histone marks for the U2OS cells WT pool, WT clones (#1 and 2), and from the 20q-TERRA KO clones (#A7, E1, E6, and C4), hybridized with a southern probe against the telomeric repeat. ChIP-dot-blot for IgG was used as a control. IgG ChIP-dot-blot shared by different antibodies shows different exposure times according to the best exposure time required for each antibody. DNA input signal is also shown. Below the ChIP-dot-blot for each mark is shown the quantification of the immunoprecipitated telomeric repeats normalized by the input for each individual sample (left graph) and for all WT clones vs. the 20q-TERRA KO clones (right graph) (mean values ± s.e.m., n = independent clone). Student's t-test was used for the statistical analysis ($^*p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$)

since the 20q-TERRA locus is an important TERRA locus but not the only one[20], we do not expect the complete abolition of TERRA expression in the 20q-TERRA KO clones. Moreover, the fact that in some of the 20q-TERRA KO clones the TERRA downregulation is only 20–50% might be related to (1) the adaptation to the cell culture conditions during clonal cell expansion and (2) the compensation of other loci, for example, the Xp[20] locus or others. Nevertheless, 20q-TERRA is the only locus in which a formal demonstration of its TERRA genuineness has been carried out by genetic means so far[20]. We confirmed TERRA's downregulation by RNA-FISH using probes to detect the TERRA's -UUAGGG-repeat. In particular, we found a significant 20–50% reduction in total TERRA levels in the 20q-TERRA KO clones compared to the controls (Fig. 1d). The 20q-TERRA KO C4 clone from our previous work[20] was included in this analysis in parallel with the newly generated clones (Fig. 1d).

Next, we set to confirm whether the newly generated 20q-TERRA KO clones also showed a telomere-shortening phenotype as previously described[20]. To this end, we performed telomeric quantitative FISH (Q-FISH) to quantify telomere fluorescence and also included formerly generated 20q-TERRA KO clone C4 as control[20]. As shown in Fig. 2, deletion of the 20q-TERRA locus



in the new clones and in the C4 clone resulted in a marked loss of telomeric sequences as seen in the switch of the distribution of telomere length frequencies toward lower values (Fig. 2a) and in the significant decrease in telomere fluorescence intensity (Fig. 2b). This was concomitant with a significant increase in the percentage of very short telomeres as determined by low telomere fluorescence (short telomeres considered those in the 10th percentile of the total telomere length distribution; Fig. 2c). We observed no differences between the pool and the WT clone #1 and an increase in telomeric length between the pool and the WT clone #2 (Fig. 2b) that could be explained as a phenomenon related to cell expansion.

In summary, the new 20q-TERRA KO clones generated here reproducibly show marked TERRA downregulation and decreased telomere length, thus supporting that the 20q locus is a bona fide TERRA origin in human U2OS cells.

TERRAs' assembly heterochromatic histone marks at telomeres. To address the role of TERRAs in telomeric chromatin formation, we evaluated the impact of 20q-TERRA deletion on the abundance of different histone marks at telomeres. To this end, we used chromatin immunoprecipitation (ChIP), followed by dot-blot (Methods). We first confirmed the presence of heterochromatic histone marks previously shown by us to be enriched at telomeres, namely tri-methylated H3K9me3 and H4K20me3[11,12] (Fig. 3a,b). Interestingly, we found that TERRA downregulation in the different 20q-TERRA KO clones resulted in a significant decrease in abundance of H3K9m3 and H4K20me3 heterochromatic marks at telomeres compared to the WT pool and WT clones (Fig. 3a,b). These findings indicate a role for TERRAs in the establishment of these telomeric heterochromatin marks. Note that the apparent lack of "dose–response" between TERRA levels and heterochromatic marks at telomeres might be related to the clonal expansion and growth adaptation of the U2OS tumoral cell line.

Next, we studied the abundance of active chromatin histone marks at telomeric chromatin, such as H3K4m3, as well as histone acetylation H3K9ac and H4K16ac marks. However, we did not find that TERRA levels significantly affected any of these active chromatin marks (Fig. 3c–e), indicating that TERRAs do not regulate active transcription histone marks at telomeres.

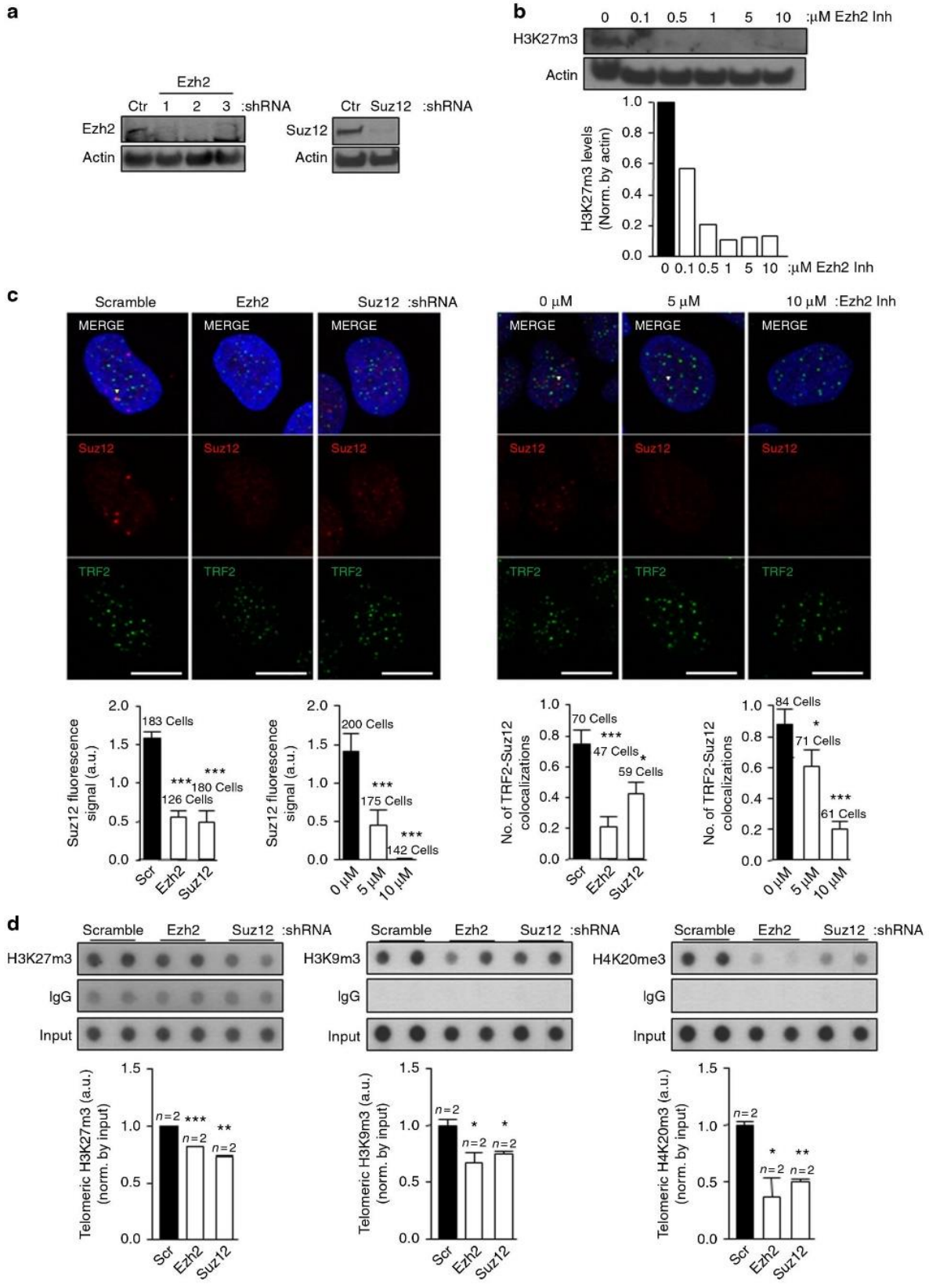
TERRAs are essential for deposition of H3K27me3 at telomeres. The facultative heterochromatin mark, H3K27me3, is established by the polycomb complex 2 and has been recently shown to cooperate with the heterochromatic histone mark H3K9me3 to recruit HP1 to heterochromatin[33]. In spite of the fact that telomeres are enriched in both HP1 and H3K9me3, a role for H3K27me3 at telomeres has not been explored before. Here, we set to address whether H3K27me3 is present at human telomeres and whether the abundance of this mark is influenced by TERRA levels. Interestingly, we found that TERRA downregulation in the different 20q-TERRA KO clones resulted in a significant decrease in abundance of H3K27me3 marks at telomeres compared to the WT pool and WT clones (Fig. 3f), demonstrating that TERRAs regulate the assembly of H3K27me3 at telomeres. The fact that all 20q-KO clones undergo similar changes at these heterochromatic marks, similar changes in TERRA levels, and have the same telomeric phenotype (see above) supports that the changes observed are not due to the accumulation of mutations during cell culture, to the monoclonal expansion, or to the presence of CRISPR off-targets. Together, these findings show that TERRAs are needed for the assembly of H3K9me3, H4K20me3, and H3K27me3 histone chromatin marks at telomeres, thus demonstrating a role for TERRAs in the establishment of telomeric heterochromatin.

TERRAs bind PRC2 and regulates its recruitment to telomeres. The finding of TERRA-dependent assembly of H3K27m3 at telomeres prompted us to investigate whether TERRAs directly regulate the presence at telomeric chromatin of the polycomb PRC2 complex responsible for catalyzing the methylation of H3K27[34]. First, we set to address a direct interaction between TERRAs and PRC2 components. To this end, we performed a TERRA biotin pull-down assay followed by western blot to detect PRC2 components. For specific TERRA binding, we incubated U2OS cells' nuclear extracts with a biotinylated oligo consisting of eight TERRA-UUAGGG repeats (8 × UUAGGG; Methods). As a negative control, we used a biotinylated oligo consisting of eight CCCUAA repeats (8 × CCUAA; the TERRA antisense sequence), as well as beads alone. To detect presence of the PRC2 complex in the pull-down we used antibodies against two of core proteins of the PRC2 complex, EZH2 and SUZ12. Importantly, we found that both EZH2 and SUZ12 were able to specifically bind to the sense TERRA oligo ("S" in Fig. 4a) but not to the control antisense oligo ("AS" in Fig. 4a). Similarly, we could not detect EZH2 and SUZ12 when only the beads were incubated with the cell extracts ("no oligo" in Fig. 4a). These findings demonstrate that the PRC2 complex specifically interacts with TERRA RNAs. During the preparation of this manuscript, the TERRA–PRC2 interaction was also proven by electrophoretic mobility shift assay using a recombinant holo-PRC2 5-mer complex (EZH2, EED, SUZ12, RBBP4, and AEBP2) and a TERRA oligo[35]. The authors also found that PRC2 has a general affinity for G-rich RNA especially those capable of folding into G-quadruplexes[35]. Later on, Chu and co-workers demonstrated the direct interaction between TERRAs and EZH2 by iDRIP (identification of direct RNA-interacting proteins)[17], which reinforces our findings.

We next set to address whether PRC2 is specifically located at telomeres and whether TERRAs are necessary for its telomeric location. To this end, we used immunofluorescence to evaluate the presence of the SUZ12 PRC2 component at telomeres. As

Fig. 4 TERRAs bind PRC2 and modulates its own and HP1 recruitment to telomeres. **a** A TERRA biotinylated RNA oligo (S) was incubated with nuclear extracts from U2OS cells and their association with Ezh2 and SUZ12 was detected by western blotting. A biotinylated control RNA oligo corresponding to the complementary sequence (AS) of the same length as the biotinylated TERRA (N_{48}) was used as control. Biotin pull-down in the absence of RNA oligo (no oligo) was included to monitor inespecific binding to the beads. **b** Representative images of the average number of colocalizations found on double immunostaining to TRF2 (green) and SUZ12 (red) in the U2OS WT and 20q-KO clones. Arrowheads indicate colocalization events. Scale bar, 10 µm. (Left graph) Quantification of the colocalization in each of the WT and 20q-TERRA KO clones (mean values ± s.e.m., n = number of cells) and (right graph) in all WT vs. the 20q-TERRA KO clones (mean values ± s.e.m., n = independent clone). **c** Telomeric ChIP-dot-blot of SUZ12 in U2OS cells infected with scramble or SUZ12 shRNA. IgG was used as a control. DNA input is also shown. (Graph) Quantification of the signal from the immunoprecipitated telomeric repeats normalized by the input (mean values ± s.e.m., n = technical triplicates). **d** Representative confocal STED super-resolution images showing the colocalization between TRF2 (in green) and SUZ12 (in red) in U2OS cells. Zoom: a colocalization event. Scale bar 5 µm. **e** Telomeric ChIP-dot-blot for HP1 in WT and 20q-TERRA KO clones. IgG was used as a control. DNA input signal is also shown. (Left graph) Quantification of the signal from the immunoprecipitated telomeric repeats normalized by the input for each individual sample and (right graph) for all WT vs. 20q-TERRA KO clones (mean values ± s.e.m., n = independent clone). Student's t-test was used for statistical analysis (*$p < 0.05$ and **$p < 0.01$)

previously described[36], we observed that SUZ12 formed nuclear foci (red dots in Fig. 4b). Interestingly, we found that approximately 1.5 SUZ12 spots (in red) per nucleus colocalized with TRF2 (in green; Fig. 4b), thus indicating the presence of the SUZ12 PRC2 protein at telomeres. If we refer this quantification to the number of telomeres, 3.5% of the telomeres (as detected by TRF2) colocalize with SUZ12. Importantly, SUZ12-TRF2 colocalization events were significantly decreased by ~50% in 20q-TERRA KO clones (#A7, #E1, #E6, and #C4) compared to the WT clones (#1 and #2), demonstrating that the SUZ12 location to telomeres was dependent on TERRAs (Fig. 4b). The specificity of this reproducible, although low-in-number interaction, of SUZ12



with the telomere was further confirmed by immunofluorescence upon downregulation of PCR2 components with shRNAs or with chemical inhibitors (see below). In addition, a SUZ12 ChIP followed by telomeric dot-blot confirmed this interaction of SUZ12 with telomeric DNA, which was decreased upon down-regulation of SUZ12 levels by shRNA (Fig. 4c). The definitive proof of the interaction of SUZ12 with the telomere was achieved by super-resolution microscopy (Fig. 4d). As seen in the confocal super-resolution images, we found colocalization events between SUZ12 and the telomeric-binding protein TRF2 (Fig. 4d). Nevertheless, additional controls to prove this interaction were performed by running three different types of randomization approaches. First, we carried out a double immunofluorescence of SUZ12 and TRF2 run in parallel with one using an anti-centromere antibody (ACA) and TRF2. As it can be seen in Supplementary Fig. 1, the number of colocalizations between SUZ12 and TRF2 was significantly higher than the ones observed by chance between TRF2 and the ACA antibody. Second, using an interaction plugin on Fiji[37], the interaction potential between TRF2-SUZ12 was calculated. Interaction is defined as the spatial distribution of signal in one channel being not independent of the signal distribution in the other channel. For that, objects were identified and their NND (nearest neighbor distance) calculated, and compared with one probability density function of the NND if the signals were independent. For the Suz-TRF2 colocalizations the strength of the interaction is superior to zero, indicating that the spatial distribution of TRF2-SUZ12 is dependant (Supplementary Fig. 2). When compared against 10,000 Monte Carlo samples of NND distributions corresponding to the null hypothesis of "no interaction", the results are statistically significant ($p < 0001$). Third, we used the randomization tool of the Definiens Developer XD.2 software. In this way, TRF2 spots were digitally found and a randomization of the TRF2 signal was performed (16,000 randomized pictures; Supplementary Fig. 3A). The number of TRF2 spots found in four original pictures and in the ones randomized is shown in the Supplementary Fig. 3B (please, note that each picture contains different number of nuclei). Next, the colocalizations between TRF2 and SUZ12 were identified both in the original and in the randomized pictures (see Methods). Importantly, we found that the number of random colocalizations was significantly lower than in the original pictures (Supplementary Fig. 3C). The percentage of randomized pictures with ≥colocalizations is significantly lower than in the original pictures (Supplementary Fig. 3D). All together, these findings demonstrate that SUZ12 interacts with the telomere and that this interaction is TERRA-dependent.

TERRAs are necessary for HP1 recruitment to telomeres. We previously showed that mammalian telomeres are enriched in HP1, a mark of constitutive heterochromatin domains[12]. Here, we set to study whether 20q-TERRA KO clones also showed altered abundance of HP1 at telomeres. Interestingly, CHIP analysis showed significantly decreased HP1 abundance at the telomeres of 20q-TERRA KO clones (#A7, #E1, #E6, and #C4) compared to the WT clones (#1 and #2), demonstrating that TERRAs are required for HP1 deposition at telomeres (Fig. 4e).

PRC2 complex is necessary for HP1 recruitment at telomeres. Since we found that TERRAs are required for H3K9me3, H3K27me3, PRC2, and HP1 recruitment at telomeres, we next set up experiments to understand mechanistically the chain of events for which TERRAs are responsible. PRC2 and H3K27me3 have been previously to cooperate with H3K9me3 tri-methylation to maintain HP1 at chromatin[33]. To assess whether this is also happening at telomeres, we first generated cells with decreased EZH2 and SUZ12 levels by using two alternative methods. In particular, we downregulated these PRC2 components using shRNAs against EZH2 or SUZ12, as well as by using the specific EZH2 chemical inhibitor EPZ-6438 (Fig. 5a,b). The down-regulation obtained with the EZH2 shRNAs was ~70–90%. the shRNA#2 being the most effective (Fig. 5a). The downregulation obtained with the SUZ12 shRNA was 85% (Fig. 5a). In the case of chemical inhibitors, we observed that the best EZH2 inhibition was achieved at concentration of 1 μM of the inhibitor with a 90% decrease in H3K27me3 global levels as determined by western blot (Fig. 5b). Next, we studied the changes at telomeres as a consequence of downregulating the PRC2 components. As expected, we found a significant decrease in global SUZ12 levels as detected by immunofluorescence in cells treated with either the EZH2 or SUZ12 shRNAs or with the EZH2 inhibitor (Fig. 5c, left bottom graphs). Consequently, the colocalization of SUZ12 with the TRF2 telomeric protein decreased significantly both with the shRNAs or upon chemical inhibition (Fig. 5c, right bottom graphs), confirming the interaction of SUZ12 with the telomere. Next, we evaluated the consequences of reduced PRC2 levels in the deposition of heterochromatic marks by telomeric ChIP. As expected because of the interaction of SUZ12 with the telomere, telomeric H3K27me3 abundance decreased significantly in cells with reduced EZH2 or SUZ12 levels (Fig. 5d, left panel). Moreover, the levels of H3K9me3 and H4K20me3 also significantly dropped in these cells (Fig. 5d, middle and right panels), indicating that PRC2 is not only important for the methylation of H3K27 but also for the deposition of these other heterochromatic marks. As a consequence of the drop in heterochromatic marks, both global HP1 levels (Fig. 6a, left bottom graphs) and its colocalization at telomeres (RAP1-HP1 colocalizations; Fig. 6a, right bottom graphs) significantly decreased both in cells treated with either the PRC2 shRNAs or upon EZH2 inhibition as observed by double immunofluorescence of HP1 and the telomeric protein RAP1. Although the presence of HP1 is a well-established mark at the telomere[12,38], we performed an additional control to prove the specificity of the HP1 antibody and of this interaction consisting of a double immunofluorescence of HP1 and the telomere-binding protein RAP1 run in parallel with one

Fig. 5 The PRC2 complex is critical for the deposition of heterochromatic histone marks and HP1 at telomeres. **a** Upon infection of U2OS cells with shRNAs against EZH2 and SUZ12, total protein was obtained and used for western blot detection of EZH2 and SUZ12. Actin was used as loading control. **b** U2OS were treated with increasing concentrations of the EZH2 inhibitor EPZ-6438 for 4 days. Nuclear protein extracts were used for western blot detection of H3K27me3. Actin was used as loading control. (Graph) Quantification is shown. **c** Representative images of the average number of colocalizations for TRF2 (green) and SUZ12 (red) in U2OS cells infected with a scramble or with EZH2 or SUZ12 shRNAs (left panel) or treated with vehicle or EZH2 inhibitor. Arrowheads indicate colocalization events. Scale bar, 10 μm. Below the images is shown the quantification of (left graphs) the total nuclear SUZ12 upon shRNAs or EZH2 inhibitor and (right graphs) the colocalization between TRF2 and SUZ12 (mean values ± s.e.m., n = number of cells). **d** Telomeric ChIP-dot-blot of the H3K27m3, H3K9m3, and H4K20m3 for U2OS cells infected with scramble, EZH2, or SUZ12 shRNAs. IgG was used as a control. IgG ChIP-dot-blot shared by different antibodies shows different exposure times according to the best exposure time required for each antibody. DNA input signal is also shown. Quantification of the immunoprecipitated telomeric repeat signal normalized by the input for each individual sample is shown below (mean values ± s.e.m., n = technical replicates). Student's t-test was used for the statistical analysis ($^*p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$)

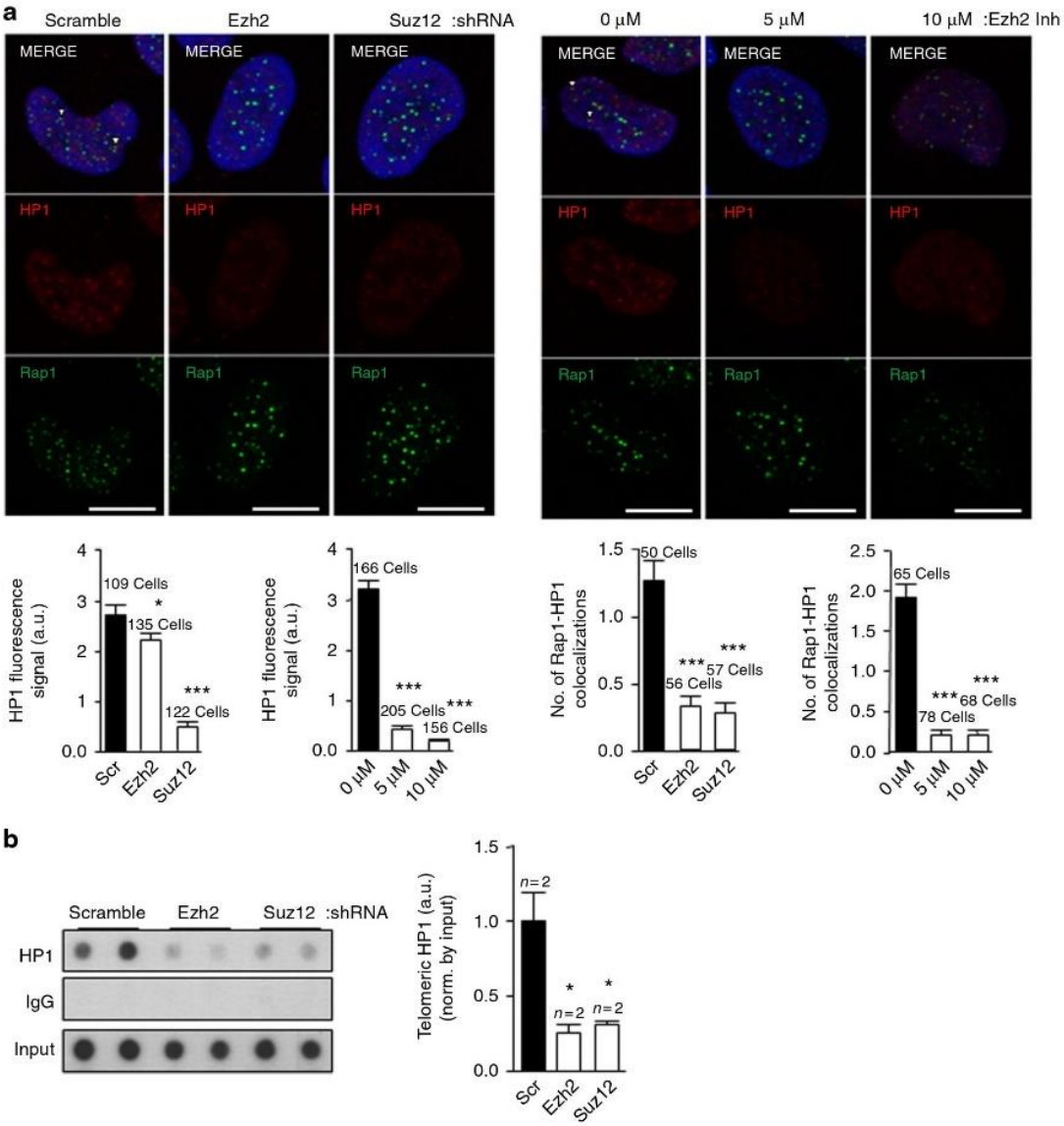


Fig. 6 Global HP1 and its localization at telomeres depend on PRC2 levels. **a** Representative images of the average number of colocalizations of RAP1 (green) and HP1 (red) in the U2OS infected with scramble, EZH2, or SUZ12 shRNA (left panel) or in U2OS treated with vehicle or with EZH2 inhibitor. Arrowheads indicate colocalization events. Scale bar, 10 μm. Quantification of (left graphs) total HP1 and (right graphs) the colocalization between RAP1 and HP1 (mean values ± s.e.m., n = number of cells). **b** Telomeric ChIP-dot-blot of the HP1 protein for the U2OS cells infected with scramble, EZH2, or SUZ12 shRNA. IgG was used as a control. DNA input signal is also shown. Quantification of the immunoprecipitated telomeric repeats normalized by the input is shown below (mean values ± s.e.m., n = technical replicates). Student's t-test was used for the statistical analysis (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$)

using ACA (antibody that recognizes the centromere) and RAP1. As can be seen in Supplementary Fig. 4, the number of colocalizations between HP1 and RAP1 was significantly higher than the ones observed by chance between RAP1 and the ACA signal. Similar to SUZ12 interaction with the telomere (see above), we performed two additional randomization approaches to prove the HP1–RAP1 interaction. First, using an interaction plugin on Fiji[37] we found that the strength of the interaction HP1–RAP1 is superior to zero, indicating that the spatial distribution of RAP1-HP1 is dependent (Supplementary Fig. 5). When compared against 10,000 Monte Carlo samples of NND distributions corresponding to the null hypothesis of "no interaction", the results are statistically significant ($p < 0.0001$). Second, using the randomization tool of the Definiens Developer XD.2 software, we first detected digitally RAP1 spots in the original and in 4.800 randomized pictures. (Supplementary Fig. 6A). The number of RAP1 spots found in four original pictures and in the ones randomized is shown in the Supplementary Fig. 6B (please, note that each picture contains different number of nuclei). Next, the colocalizations between RAP1 and HP1 were identified both in the original and in the randomized pictures (see Methods). Importantly, we found that the number of random colocalizations was significantly lower than in the original pictures (Supplementary Fig. 6C). The percentage of randomized pictures with ≥colocalizations is significantly lower than in the original pictures ($p > 0.01$; Supplementary Fig. 6D). Moreover, we confirmed the significant decrease in HP1 abundance at telomeres by telomeric ChIP in cells with reduced EZH2 or SUZ12 levels (Fig. 6b). Note that HP1 levels do not change upon downregulation of either EZH2 or SUZ12 (Supplementary Fig. 7)

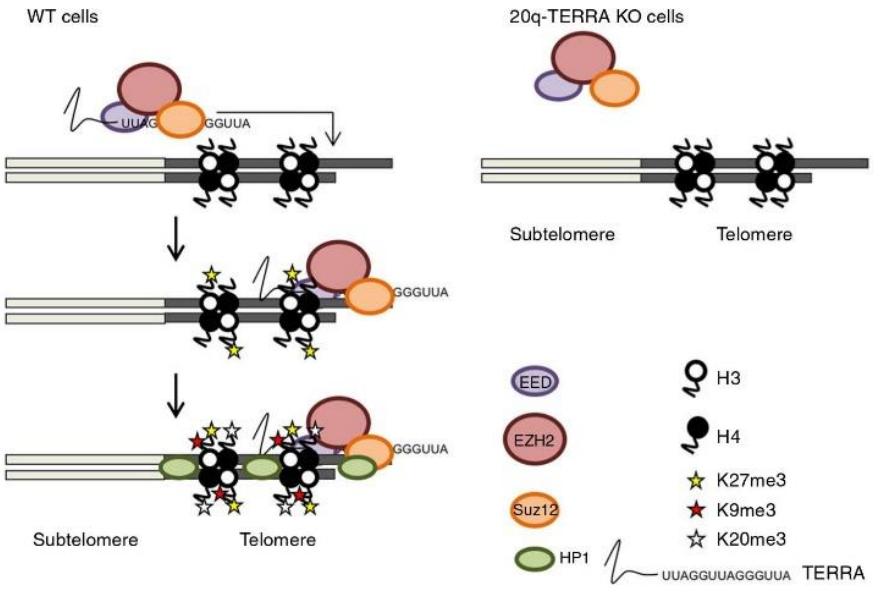


Fig. 7 Model of TERRAs as a master regulator of the heterochromatic status of the telomere. Diagram showing how TERRA recruits the PRC2 complex (EED, EZH2, SUZ12) and directs it to the telomere. Upon binding of PRC2 to the telomere, PRC2 catalyzes H3K27 methylation. This mark facilitates then the deposition of H3K9m3 and H4K20m3 and the recruitment and stabilization of HP1 protein at the telomere, important to maintain the heterochromatic status of the telomere. In the 20q-TERRA KO cells PRC2 is not recruited to the telomere, and the heterochomatic marks and HP1 protein are not taking place

Together, these results are in agreement with a model in which TERRAs interact with the PRC2 complex, and this interaction is important for PRC2 deposition to telomeres. In turn, this deposition is critical for the establishment of the H3K27me3 histone mark at telomeres, as well as for the subsequent assembly of H3K9me3 and H4K20m34 heterochromatin marks. Upon the establishment of these marks, HP1 is recruited to telomeres (Fig. 7).

TERRAs do not affect subtelomeric and total DNA methylation. Given the role of TERRAs in heterochromatin assembly at telomeric chromatin, we next studied whether TERRAs can also influence subtelomeric DNA methylation as well as global DNA methylation. Subtelomeric DNA is highly methylated at CpG-rich regions[10,11], being the methylation of the D4Z4 repeat representative of the methylation state of the subtelomere. For this reason, we performed bisulfite pyrosequencing to evaluate the DNA methylation levels at the D4Z4 repeats in our panel of cells WT and 20q-TERRA KO. We did not observe differences in the different CpG analyzed (cytosines C1, 2, and 3) between the 20q-TERRA KO clones and the WT clones (Supplementary Fig. 8). The slight differences found between clones of the same genotype might be related to inner differences between the clones (Supplementary Fig. 8A). When we calculated the percentage of changes in methylated CpG of the 20q-TERRA KO clones with respect to the control, we did not find any statistical differences between the WT clones and the 20q-TERRA KO clones, in any of the cytosine evaluated (Supplementary Fig. 8B).

To assess the role of TERRAs in total DNA methylation, we performed bisulfite pyrosequencing on LINE1 repeats, which are representative of the global DNA methylation state. As with the subtelomeric DNA methylation, we did not find differences between the WT and the 20q-TERRA KO clones (Supplementary Fig. 8C). When we calculated the percent increase of methylated CpG of the 20q-TERRA KO clones with respect to the WT pool, we did not find significant differences either (Supplementary Fig. 8D). CpG methylation at Alu repeats was used as control (Supplementary Fig. 8E).

The above data indicate that TERRAs do not have a clear role on subtelomeric and total DNA methylation.

Discussion

Telomeres are known to be enriched in heterochromatin histone tri-methylation marks, namely H3K9me3 and H4K20me3, as well as HP1 binding[12,39,40], thus being part of the so-called constitutive heterochromatin. In addition, subtelomeric DNA is hypermethylated[7,11,13]. Both histone and DNA hypermethylation marks contribute to the repressive environment at telomeres, acting as negative regulators of telomere length and telomere recombination[8–13]. Interestingly, TERRAs are known to bind both in *cis* and in *trans* to telomeric chromatin[15,16,18]; however, the roles of TERRAs at the assembly of telomeric chromatin are still vaguely defined owing to lack of TERRA knockout models, although an interaction between TERRAs and H3K9me3 and HP1 has been described before[21].

We recently showed that the majority of TERRAs in human U2OS ALT cells arise from the 20q-TERRA locus[20]. We have also generated human U2OS cells knockout for this 20q-TERRA locus, and showed that this resulted in markedly decreased global TERRA levels, as well as loss of telomere protection and telomere shortening[20]. Here, we used this TERRA knockout cellular model to address the role of TERRAs in the establishment of telomeric chromatin. We found that human U2OS cells KO for 20q-TERRA show decreased abundance of the heterochromatin H3K9me3 and H4K20m3 histone tri-methylation marks at telomeres, previously shown by us to be important for the repressive environment of telomeres[12,39,40]. In addition, we found that DNA hypermethylation of subtelomeric sequences, which is another feature of telomeric chromatin[11], was not affected by TERRA depletion, indicating that this mark is independent of TERRA levels. Therefore, our results demonstrate that TERRAs are necessary for the establishment of the bona fide H3K9me3 and H4K20m3 heterochromatic histone marks at telomeres but not for subtelomeric DNA hypermethylation.

Interestingly, we found that 20q-TERRA KO cells also showed a decreased abundance of the facultative heterochromatin mark

H3K27me3 at telomeres, which has only been described in lower eukaryotes such as unicellular algae[41]. In the filamentous fungus *Neurospora crassa* the H3K27me2/3 mark is principally sub-telomeric covering ≈7% of the genome, and its loss results in telomere mislocalization and alterations in normal chromosome conformation[42]. As this mark is established by the Polycomb Repressive Complex 2 (PRC2) we next seek whether there was a direct interaction between TERRAs and the PRC2 complex. We found that TERRAs directly bound to the PRC2 components EZH2 and SUZ12. Furthermore, we saw that SUZ12 specifically colocalized with telomeric chromatin by different techniques, including immunoprecipitation of SUZ12 followed by telomeric dot-blot and, more importantly, by colocalization of SUZ12 with the telomeric binding protein TRF2, using super-resolution confocal microscopy. Interestingly, this interaction was decreased in the absence of TERRAs. These findings are in agreement with a recent report indicating that TERRA location to extratelomeric sites influences H3K27 deposition at these sites through interaction with the PRC2[17].

Our data support a role of TERRAs as key factors in the deposition of heterochromatic marks at telomeres in ALT human cells. It was shown before that TERRA binds H3K9me3 and HP1 (ref. [21]) but here we have completed the picture with the finding of PRC2 being the TERRA mediator for the heterochromatization of human telomeres. Thus, we found that TERRAs interact with PRC2 and that PRC2 depletion leads to loss of heterochromatic marks at telomeres. Interestingly, genome-wide targeting of PRC2 depends on ATRX[43]. However, U2OS cells lack ATRX and, therefore, it is quite possible that both PRC2 localization to telomeres and the interplay between facultative and heterochromatic marks at telomeres in U2OS cells is very different from those in telomerase-positive cells. Moreover, ALT cells are defective in the cell cycle regulation of TERRA and neither the levels of TERRA nor the colocalization of TERRA at telomeres declined from S to G_2 as it occurs in telomerase-positive cells[22]. This persistent TERRA presence at telomeres might be important for PRC2 recruitment in the absence of ATRX. On the other hand, the binding of TERRAs to PRC2 is conserved in mammals as it was also identified to occur in mouse embryonic stem cells[17]. In this murine cells, TERRA binding was strongly correlated with H3K27me3 genome wide[17], supporting our finding of TERRA regulating H3K27me3 levels in human telomeres. We went one step further and found that TERRAs do not only regulate H3K27me3 deposition but also H3K9me3, H4K20me3 and HP1 and that this is mediated by PRC2. According to the role described for PRC2 and H3K27me3 cooperating with H3K9 methylation to maintain HP1 at chromatin[33], our findings at telomeres are also in agreement with this heterochromatization process. Moreover, our data indicate that TERRAs would be the key molecule to lead this process. Whether this important role could be extensive to telomerase-positive cells awaits to be tested. In summary, we describe for the first time a role for PRC2 in the establishment of telomeric chromatin, which is regulated by TERRAs. In addition, we demonstrate an important role for TERRAs in telomeric heterochromatin assembly.

Methods

Cells, transfection, infection and treatments. Human U2OS (ATCC) were cultured according to the ATCC's recommendations. Plasmids were electroporated using the Neon Transfection System (Invitrogen) following the manufacturers' protocol. U2OS cells were infected with an shRNA against the PRC2 complex subunits, which were cloned into Plko.1-Puro vector (shRNA sequence available in Supplementary Table 1). Lentiviruses were packaged in 293T cells (ATCC-CRL-3216) using the third-generation packaging system vectors, pMDLg/pRRE, pRSV. Rev, pMDG VSVG. Cells were seeded at a 50% of confluency 24 h before infection. Two infections were performed every 24 h by adding 3 ml of viral supernatant. Then, cells were allowed to recover for 24 h in growth medium before undergoing selection with puromycin for 2 days. For EZH2 inhibition U2OS cells were cultured with DMEM that contained different concentrations of the EZH2 inhibitor EPZ-6438 (SELLECKCHEM) for 4 days.

Generation of TERRA KO clones using the CRISPR-Cas9 system. The 20q-TERRA KO clones were generated using an all-in-one plasmid that contains the two gRNAs (S2 and E1) needed for the deletion of the 20q-TERRA locus. The backbone of the all-in-one plasmid was the pSpCas9(BB)-2A-GFP that already contains the S2 gRNA[20]. The gRNA End1 (E1) with the U6 promoter was amplified by PCR adding Not1 restriction sequence (primers available in Supplementary Table 1) with the GoTaq polymerase (Promega) using as template the pSpCas9(BB)-2A-GFP that already contains the E1 gRNA[20]. The PCR product and the pSpCas9(BB)-2A-GFP containing the Start 2 (S2) gRNA[20] were digested with the Not1 enzyme (New England Biolabs) during 4 h at 37 °C, and after ligation was performed at 16 °C overnight with the T4 DNA ligase enzyme (New England Biolabs). The plasmid with the new insert was transformed in One Shot TOP10 Chemically Competent E. coli DH5α (Invitrogen) and plated on LB agar plates with penicillin. The plasmid was purified with Plasmid Miniprep Kit (QIAGEN) and the presence of the insert confirmed with Sanger-style BigDye terminator chemistry on an ABI 3730 × l sequencer (Applied Biosystems). Then, the all-in-one plasmid containing the Cas9, both gRNA (S2 and E1), and a GFP cassette was transiently electroporated into the cells using the Neon system. Two days after transfection cells were trypsinized, washed with dPBS, and three or five cells of the 10% GFP brightest ones were sorted using the FACS ARIA IIU (Becton Dickinson) and plated into 96-well plates. Two weeks later, wells were checked for monoclonal cell expansion and those clones were selected for genotyping by PCR. Homozygous monoclonal cell lines were expanded. PCR primers used for genotyping can be found in Supplementary Table 1.

RNA dot-blot. RNA dot-blot analyses were performed using standard protocols. TERRA probe was obtained from a 1.6-kb $(TTAGGG)_n$ cDNA insert excised from pNYH3 (kind gift from T. de Lange, Rockefeller University, NY, USA). Dot-blot was normalized using 18S probes and quantified using ImageJ. The probes were labeled using the commercial Prime-It II Random Primer Labeling Kit (Agilent Genomics).

RNA-FISH. Cells grown on poly-L-lysine-coated coverslips (Becton Dickinson) were placed in cytobuffer (100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 10 mM pipes pH 6.8) for 30 s, washed in cytobuffer with 0.5% Triton X-100 for 30 s, washed in cytobuffer for 30 s, and then fixed for 10 min in 4% paraformaldehyde in phosphate-buffered saline (PBS). The cells were dehydrated in 70, 80, 95, and 100% ethanol, air-dried, and hybridized overnight at 37 °C with a telomere-specific PNA-FITC probe (Panagene) in hybridization buffer (2 × sodium saline citrate (SSC)/ 50% formamide). Coverslips were washed two times for 15 min in hybridization buffer at 40 °C, two times for 10 min in 2 × SSC at 40 °C, 10 min in 1 × SSC at 40 °C, 5 min in 4 × SSC at room temperature, 5 min in 4 × SSC containing 0.1% Tween-20 and DAPI (Molecular Probes) at room temperature, and 5 min in 4 × SSC at room temperature. Signals were visualized in a confocal ultraspectral microscope SP5-WLL (Leica). TERRA signal was quantified using the Definiens Developer XD.2 software.

High-throughput telomere length quantification by FISH. HT-q-FISH was done according to ref. [44]. Cells were plated into clear-bottom black-walled 96-well plates pre-coated for 30 min with 0.1% porcine gelatin. Then, cells were grown in DMEM at 37 °C overnight, fixed in methanol/acetic acid (3:1, *v/v*) two times for 10 min, and were left in the fixative overnight at −20 °C. Fixative was removed, plates dried for at least 1 h at 37 °C, and samples were rehydrated in PBS. Plates were then subjected to a standard Q-FISH protocol using a telomere-specific PNA-CY3 probe; DAPI was used to stain nuclei. Images were captured using the OPERA (Perkin Elmer) High-Content Screening system. TL values were analyzed using individual telomere spots. Samples were analyzed in triplicate.

Biotin pull-down analysis. Biotin pull-down assays were carried out as described in ref. [45] except for that 150 µg of nuclear lysate were incubated with 0.9 ng of biotinylated transcripts (Invitrogen) for 1 h at room temperature. An amount of 0.9 ng of total RNA was added as competitor. Complexes were isolated using streptavidin-conjugated Dynabeads (Dynal), and bound proteins in the pull-down material were analyzed by western blotting. The biotinylated TERRA sense transcript consists of 8 × UUAGGG; the control biotinylated transcripts consist of the TERRA antisense sequence 8 × CCCUAAA. See antibodies in the western blot section. Full gels can be found in Supplementary Fig. 9.

Western blot analysis. Whole-cell lysates and nuclear lysates were prepared using RIPA buffer and RSB buffer (10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 3 mM $MgCl_2$, and inhibitors), respectively, as previously described[19]. Protein lysates were resolved by SDS–PAGE and transferred onto nitrocellulose membranes. Antibodies used were the following: EZH2 (D2C9; Cell signaling), SUZ12 (ab12073; Abcam), H3K27m3 (07–449; Millipore), HP1γ (clone 42S2; Millipore), and ACTIN (A5441;

Sigma-Aldrich). Following secondary antibody incubations, signals were visualized by enhanced chemiluminescence. Full gels can be found in Supplementary Fig. 9.

Immunofluorescence and super-resolution microscopy. Cells grown on poly-L-lysine-coated coverslips (Becton Dickinson) were placed in cytobuffer (100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 10 mM pipes pH 6.8) with 0.5% Triton X-100 for 6 min and then fixed for 10 min in 4% paraformaldehyde in PBS. Cells were blocked with 10% BSA in PBS for 1 h at room temperature. Cells were incubated with primary antibody dissolved in Dako antibody Diluent (Dako) for 1 h in a humid chamber at room temperature. Coverslips were washed three times for 30 min in PBS containing 0.1% Tween-20. Cells were incubated with Alexa secondary antibody (Life Technologies, A11017) dissolved in Dako antibody Diluent (Dako) for 1 h in a humid chamber at room temperature. Cells were washed three times for 30 min in PBS containing 0.1% Tween-20. Samples were mounted in Prolong with DAPI (Invitrogen). Signals were visualized in a confocal ultraspectral microscope SP5-WLL (Leica). The following antibodies were used: TRF2 (clone 4A794; Millipore) diluted 1/250, SUZ12 (ab12073; Abcam) diluted 1/500, RAP1 (A300–306A; Bethy) diluted 1/250, HP1γ (clone 42S2; Millipore) diluted 1/200, and Anti Centromere Antibody (ACA; 15–235; Antibodies Incorporated) diluted 1/100. For Super-resolution microscopy acquisition, we use a confocal multispectral Leica TCS SP8 system with a 3 × STED module for super-resolution. Laser lines: 405 nm and WLL2 (white laser for 470–670 nm excitation). Depletion lines: 592 and 660 nm.

Randomization analysis. Two different randomization analyses were performed. First, the MosacIA Fiji plugin (https://imagej.net/Interaction_Analysis)[37,46] was used to calculate the interaction potentials. Potential used was Plummer (100 iterations). Grid spacing was 0.5 pixels. Kernel wt(q) was 0.001. Kernel wt(p) was provided by the plugin using Silverman's rule. The results were tested for significance against 10,000 Monte Carlo samples of point distributions corresponding to the null hypothesis of "no interaction"[47]. Second, the randomization tool of the Definiens Developer XD.2 software was used. TRF2 spots were found in DNA-positive nuclear areas. Those regions without DNA signal, where TRF2 is not found but SUZ12 is detected (as in the nucleoli), were eliminated from the analysis. Colocalizations between TRF2 and SUZ12 were identified according to SUZ12 signal intensity and a colocalization of TRF2 signal on SUZ12 of more than 60%. The number of colocalizations between SUZ12 and the virtual TRF2 spots were calculated using the same criteria for the original and randomized images. In all, 16,000 randomized pictures were generated. In a similar manner, the colocalizations of RAP1-HP1 were calculated in the original pictures and in 4.800 randomized pictures. The RAP1 spots were detected in the entire nucleus.

ChIP assay and telomere dot-blots. For ChIP analysis, we used 4×10^6 cells per condition. Formaldehyde was added directly to tissue culture medium to a final concentration of 1% and left for 10 min at room temperature on a shaking platform. The crosslinking was stopped by adding glycine to a final concentration of 0.125 M. Crosslinked cells were washed twice with cold PBS, scraped and lysed at a density of 5×10^6 cells ml^{-1} for 10 min at 4°C in 1% SDS, 50 mM Tris-HCl (pH 8.0) and 10 mM EDTA containing protease inhibitors. Lysates were sonicated to obtain chromatin fragments <1 kb and centrifuged for 15 min in a microfuge at room temperature. An aliquot of 200 µl of lysate was diluted 1:10 with 1.1% Triton X-100, 2 mM EDTA, 150 mM NaCl, and 20 mM Tris-HCl (pH 8.0) containing protease inhibitors, and precleared with 50 µl of Protein A/G Plus-Agarose sc-2003 beads (Santa Cruz Biotechnology). After centrifugation, the fragments were incubated with 4 µg of the following antibodies: H3K27m3 (07–449, Millipore), H3K9m3 (07–442, Upstate), H4K20m3 (07–749, Upstate), H3K4m3 (CS200580, Millipore), H3K9ac (072K4824, Sigma-Aldrich), H4K16ac (cat39167, Active Motif), or normal rabbit IgG (Santa Cruz Biotechnology) and 10 µg of: HP1γ (clone 42S2; Millipore) and SUZ12 (ab12073; Abcam) diluted at 4 °C overnight on a rotating platform. Next day, 50 µl of Protein A/G Plus-Agarose (sc-2003) beads were added and incubated for 1 h. The immunoprecipitated pellets were washed once with 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl (pH 8.0), and 150 mM NaCl, and then with 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl (pH 8.0), and 500 mM NaCl, and next with 0.25 M LiCl, 1% Nonidet P-40, 1% sodium deoxycholate, 1 mM EDTA, and 10 mM Tris-HCl, pH 8.0 and finally with 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA two times. The chromatin was eluted from the beads twice by incubation with 250 µl 1% SDS and 0.1 M $NaHCO_3$ during 15 min at room temperature with rotation. After adding 20 µl of 5 M NaCl, the crosslink was reversed overnight at 65 °C. Samples were supplemented with 20 µl of 1 M Tris-HCl (pH 6.5), 10 µl of 0.5 M EDTA, 20 µg of RNase A, and 40 µg of proteinase K, and were incubated for 1 h at 45 °C. DNA was recovered by phenol–chloroform extraction and ethanol precipitation, slot-blotted it onto a Hybond N+ membrane, and hybridized it with a plasmid containing 1.6 Kb of TTAGGG repeats (gift from T. de Lange, Rockefeller University, USA). The signal was quantified with the ImageJ software. For total telomeric DNA samples, 500 µl of lysate was processed with the rest of the samples at the step of reversing the crosslinks. The amount of telomeric DNA immunoprecipitated was calculated in each ChIP based on the signal relative to the corresponding total telomeric DNA signal. Full blots can be found in Supplementary Fig. 9.

Data availability. All relevant data are available from the authors.

Received: 12 October 2017 Accepted: 20 March 2018
Published online: 18 April 2018

References

1. Chan, S. W. & Blackburn, E. H. New ways not to make ends meet: telomerase, DNA damage proteins and heterochromatin. *Oncogene* **21**, 553–563 (2002).
2. de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
3. Funk, W. D. et al. Telomerase expression restores dermal integrity to in vitro-aged fibroblasts in a reconstituted skin model. *Exp. Cell Res.* **258**, 270–278 (2000).
4. Blasco, M. A., Funk, W., Villeponteau, B. & Greider, C. W. Functional characterization and developmental regulation of mouse telomerase RNA. *Science* **269**, 1267–1270 (1995).
5. Shay, J. W. & Wright, W. E. Role of telomeres and telomerase in cancer. *Semin. Cancer Biol.* **21**, 349–353 (2011).
6. Dunham, M. A., Neumann, A. A., Fasching, C. L. & Reddel, R. R. Telomere maintenance by recombination in human cells. *Nat. Genet.* **26**, 447–450 (2000).
7. Blasco, M. A. The epigenetic regulation of mammalian telomeres. *Nat. Rev. Genet.* **8**, 299–309 (2007).
8. Benetti, R., Garcia-Cao, M. & Blasco, M. A. Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. *Nat. Genet.* **39**, 243–250 (2007).
9. Benetti, R. et al. Suv4-20h deficiency results in telomere elongation and derepression of telomere recombination. *J. Cell Biol.* **178**, 925–936 (2007).
10. Brock, G. J., Charlton, J. & Bird, A. Densely methylated sequences that are preferentially localized at telomere-proximal regions of human chromosomes. *Gene* **240**, 269–277 (1999).
11. Gonzalo, S. et al. DNA methyltransferases control telomere length and telomere recombination in mammalian cells. *Nat. Cell Biol.* **8**, 416–424 (2006).
12. Garcia-Cao, M., O'Sullivan, R., Peters, A. H., Jenuwein, T. & Blasco, M. A. Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nat. Genet.* **36**, 94–99 (2004).
13. Schoeftner, S. & Blasco, M. A. A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. *EMBO J.* **28**, 2323–2336 (2009).
14. Marion, R. M. et al. Telomeres acquire embryonic stem cell characteristics in induced pluripotent stem cells. *Cell Stem. Cell.* **4**, 141–154 (2009).
15. Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**, 798–801 (2007).
16. Schoeftner, S. & Blasco, M. A. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat. Cell Biol.* **10**, 228–236 (2008).
17. Chu, H. P. et al. TERRA RNA antagonizes ATRX and protects telomeres. *Cell* **170**, 86–101 e116 (2017).
18. Lopez de Silanes, I. et al. Identification of TERRA locus unveils a telomere protection role through association to nearly all chromosomes. *Nat. Commun.* **5**, 4723 (2014).
19. Lopez de Silanes, I., Stagno d'Alcontres, M. & Blasco, M. A. TERRA transcripts are bound by a complex array of RNA-binding proteins. *Nat. Commun.* **1**, 33 (2010).
20. Montero, J. J., Lopez de Silanes, I., Grana, O. & Blasco, M. A. Telomeric RNAs are essential to maintain telomeres. *Nat. Commun.* **7**, 12534 (2016).
21. Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. & Lieberman, P. M. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol. Cell* **35**, 403–413 (2009).
22. Flynn, R. L. et al. Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. *Science* **347**, 273–277 (2015).
23. Flynn, R. L. et al. TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature* **471**, 532–536 (2011).
24. Graf, M. et al. Telomere length determines TERRA and R-loop regulation through the cell cycle. *Cell* **170**, 72–85 e14 (2017).
25. Balk, B. et al. Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nat. Struct. Mol. Biol.* **20**, 1199–1205 (2013).
26. Arora, R. et al. RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat. Commun.* **5**, 5220 (2014).
27. Rinn, J. L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
28. Bernstein, E. & Allis, C. D. RNA meets chromatin. *Genes Dev.* **19**, 1635–1655 (2005).

29. Muller, S. & Almouzni, G. Chromatin dynamics during the cell cycle at centromeres. *Nat. Rev. Genet.* **18**, 192–208 (2017).
30. Castellano-Pozo, M. et al. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol. Cell* **52**, 583–590 (2013).
31. Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**, 436–439 (2014).
32. Torres-Ruiz, R. et al. Efficient recreation of t(11;22) EWSR1-FLI1 + in human stem cells using CRISPR/Cas9. *Stem Cell Rep.* **8**, 1408–1420 (2017).
33. Boros, J., Arnoult, N., Stroobant, V., Collet, J. F. & Decottignies, A. Polycomb repressive complex 2 and H3K27me3 cooperate with H3K9 methylation to maintain heterochromatin protein 1alpha at chromatin. *Mol. Cell Biol.* **34**, 3662–3674 (2014).
34. Muller, J. et al. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* **111**, 197–208 (2002).
35. Wang, X. et al. Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. *Mol. Cell* **65**, 1056–1067 e1055 (2017).
36. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
37. Shivanandan, A., Radenovic, A. & Sbalzarini, I. F. MosaicIA: an ImageJ/Fiji plugin for spatial pattern and interaction analysis. *BMC Bioinformatics* **14**, 349 (2013).
38. Canudas, S. et al. A role for heterochromatin protein 1gamma at human telomeres. *Genes Dev.* **25**, 1807–1819 (2011).
39. Garcia-Cao, M., Gonzalo, S., Dean, D. & Blasco, M. A. A role for the Rb family of proteins in controlling telomere length. *Nat. Genet.* **32**, 415–419 (2002).
40. Gonzalo, S. et al. Role of the RB1 family in stabilizing histone methylation at constitutive heterochromatin. *Nat. Cell Biol.* **7**, 420–428 (2005).
41. Mikulski, P., Komarynets, O., Fachinelli, F., Weber, A. P. M. & Schubert, D. Characterization of the polycomb-group mark H3K27me3 in unicellular algae. *Front. Plant Sci.* **8**, 607 (2017).
42. Klocko, A. D. et al. Normal chromosome conformation depends on subtelomeric facultative heterochromatin in *Neurospora crassa*. *Proc. Natl Acad. Sci. USA* **113**, 15048–15053 (2016).
43. Sarma, K. et al. ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. *Cell* **159**, 869–883 (2014).
44. Canela, A., Klatt, P. & Blasco, M. A. Telomere length analysis. *Methods Mol. Biol.* **371**, 45–72 (2007).
45. Wang, W. et al. HuR regulates p21 mRNA stabilization by UV light. *Mol. Cell Biol.* **20**, 760–769 (2000).
46. Helmuth, J. A., Paul, G. & Sbalzarini, I. F. Beyond colocalization: inferring spatial interactions between sub-cellular structures from microscopy images. *BMC Bioinformatics* **11**, 372 (2010).
47. Silverman, B. W. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability book series (Chapman and Hall, London, 1986).

Acknowledgements

We thank to R. Torres and S. Rodriguez for advice in the CRISPR-Cas9 technology and helpful discussions. We are indebted to Sylvia Gutiérrez Erlandsson, principal investigator of the confocal microscopy Group at the National Centre of Biotechnology, Madrid (Spain), because, without her generous help, we could not have performed the super-resolution microscopy. Research in the Blasco lab is funded by the Spanish Ministry of Economy and Competitiveness Projects (SAF2013-45111-R and SAF2015-72455-EXP), the World Cancer Research (WCR) Project (16-1177), and the Fundación Botín (Spain).

Author contribution

M.A.B. conceived the original idea. J.J.M. conducted and analyzed the experiments. J.J. M., I.I.S., and M.A.B. designed the experiments, D.M. did the super-resolution microscopy acquisition and performed the immunofluorescence image quantification and the Definiens randomization. A.C.G. did the MosaicIA randomization analysis. M.F.F. performed the DNA methylation analysis; J.J.M., I.I.S., and M.A.B. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-03916-3.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Artículo 3

TERRA regulate the transcriptional landscape of pluripotent cells throught TRF1-dependent recruitment of PRC2

Artículo 3: TERRA regulate the transcriptional landscape of pluripotent cells throught TRF1-dependent recruitment of PRC2.

Autores: Rosa M. Marión*, **Juan José Montero*,** Isabel López-Silanes, Osvaldo Graña, Paula Martínez, Stefan Schoeftner, Jose A. Palacios-Fabrega y María A. Blasco
(* Estos autores han contribuido de forma igualitaria al manuscrito).

Publicado en Elife. 20 de agosto de 2019. 8, identificador de artículo: e44656

Resumen:

La proteína TRF1 forma parte complejo shelterina o telosoma. TRF1 esencial para la protección telomérica, al evitar que los telómeros sean reconocidos como roturas de doble cadena en el DNA evitando que se fusionen y degraden. Además, TRF1 es importante en el mantenimiento de la pluripotencia. En este sentido TRF1 es un marcador de células madre y necesaria para el mantenimiento de las células madre adultas y pluripotentes, además de para el proceso el proceso de reprogramación. Pero hasta ahora, el mecanismo por el cual TRF1 es esencial para las células madre y la pluripotencia era desconocido.

Con el fin de entender el papel de TRF1 en pluripotencia, hemos usado shRNAs contra TRF1 en IPSCs (células madre pluripotentes inducidas) en estado naïve y KO para P53. Tras el silenciamiento de TRF1, usamos RNA-seq para estudiar los cambios de expresión dependientes de TRF1. Observamos cambios en la expresión de cientos de genes, la mayoría relacionados con el manteamiento de la pluripotencia y diferenciación celular. Un gran porcentaje de estos son genes regulados por el complejo PRC2. Para comprobar si este complejo estaba relacionado con los cambios transcripcionales, realizamos ChIP-seq contra PRC2 y su marca H3K27me3 en células control y células en las que silenciamos TRF1. De esta forma demostramos, que la eliminación de TRF1 en IPSCs, causa un aumento de la unión por todo el genoma de PRC2 y un aumento en la deposición de H3K27me3. Además, esta unión ocurre también en la mayoría de los promotores de los genes cuya expresión cambia al eliminar TRF1. Además, el silenciamiento de TRF1 causa un incremento en la transcripción de TERRA. Este incremento de la expresión de TERRA está asociado con una mayor unión de TERRA a los promotores de genes cuya expresión cambia al quitar TRF1.

En conclusión, estos datos son consistentes con un modelo por el cual TRF1 es esencial para el mantenimiento de la pluripotencia al regular los niveles de TERRA. Al bajar los niveles de TRF1, aumentarían los de TERRA, TERRA se uniría a los promotores de genes implicados en pluripotencia y diferenciación, reclutando de esa forma al complejo PRC2 y alterando su expresión.

Artículo 3

Contribución Personal: He participado en el diseño del estudio, realizado y analizado aproximadamente la mitad de los experimentos (análisis de enriquecimiento de los diferentes conjuntos de genes, clonaje de shRNAs y silenciamiento en IPSCs, medición de los niveles de TERRA, CHIRT-seq). Finalmente, he contribuido en la discusión de los resultados y en la preparación del artículo al elaborar parte de las figuras.

eLIFE elifesciences.org

RESEARCH ARTICLE

TERRA regulate the transcriptional landscape of pluripotent cells through TRF1-dependent recruitment of PRC2

Rosa María Marión[1†], Juan J Montero[1†], Isabel López de Silanes[1], Osvaldo Graña-Castro[2], Paula Martínez[1], Stefan Schoeftner[1‡§], José Alejandro Palacios-Fábrega[1#], Maria A Blasco[1*]

[1]Telomeres and Telomerase Group, Molecular Oncology Program, Spanish National Cancer Centre (CNIO), Madrid, Spain; [2]Bioinformatics Unit, Structural Biology Program, Spanish National Cancer Centre (CNIO), Madrid, Spain

Abstract The mechanisms that regulate pluripotency are still largely unknown. Here, we show that Telomere Repeat Binding Factor 1 (TRF1), a component of the shelterin complex, regulates the genome-wide binding of polycomb and polycomb H3K27me3 repressive marks to pluripotency genes, thereby exerting vast epigenetic changes that contribute to the maintenance of mouse ES cells in a naïve state. We further show that TRF1 mediates these effects by regulating TERRA, the lncRNAs transcribed from telomeres. We find that TERRAs are enriched at polycomb and stem cell genes in pluripotent cells and that TRF1 abrogation results in increased TERRA levels and in higher TERRA binding to those genes, coincidental with the induction of cell-fate programs and the loss of the naïve state. These results are consistent with a model in which TRF1-dependent changes in TERRA levels modulate polycomb recruitment to pluripotency and differentiation genes. These unprecedented findings explain why TRF1 is essential for the induction and maintenance of pluripotency.

DOI: https://doi.org/10.7554/eLife.44656.001

***For correspondence:** mblasco@cnio.es

[†]These authors contributed equally to this work

Present address: [‡]Genomic Stability Unit, Laboratorio Nazionale del Consorzio Interuniversitario per le Biotecnologie (LNCIB), Trieste, Italy; [§]Department of Life Sciences, Università degli Studi di Trieste, Trieste, Italy; [#]Astellas Pharma Europe Ltd, Chertsey, United Kingdom

Competing interest: See page 28

Funding: See page 28

Received: 21 December 2018
Accepted: 26 July 2019
Published: 20 August 2019

Reviewing editor: Jeannie T Lee, Massachusetts General Hospital, United States

Introduction

Multiple cellular processes, including pluripotency and the determination of cell fate, are regulated by epigenetic modifications, including genome-wide changes in DNA and histone methylation (***Blasco, 2007***; ***Theunissen and Jaenisch, 2014***). One key histone modification, tri-methylation of lysine 27 in histone 3 (H3K27me3), is a transcription repression mark that it is controlled by the polycomb transcriptional repressor (PRC) proteins (***Sparmann and van Lohuizen, 2006***). There are two polycomb complexes, PRC1 and PRC2, which cooperate to achieve gene silencing (***Lund and van Lohuizen, 2004***). PRC2 encompasses EED (Embryonic ectoderm development), EZH2 (Enhancer of zeste), SUZ12 (Suppressor of zeste 12), and ESC (Extra sex combs) (***Ringrose and Paro, 2004***), and is involved in the initiation of gene repression. In particular, the PRC2 component EZH2 is a SET-domain-containing protein that catalyzes the repressive H3K27me3 mark and to lesser extent also H3K9me3 (***Simon and Kingston, 2009***). PRC2-mediated H3K27me3 recruits PRC1, which contributes to gene silencing, possibly by blocking transcriptional elongation by the RNA Polymerase II (***Stock et al., 2007***).

Interestingly, polycomb PRC1 and PRC2 complexes have been proposed to have a key role in the maintenance of embryonic stem (ES) cell pluripotency (***Pereira et al., 2010***). In particular, polycomb proteins are important for restraining the activity of lineage-specifying factors in ES cells, and also have been shown to be essential in establishing the conversion of differentiated cells towards pluripotency (***Azuara et al., 2006***; ***Boyer et al., 2006***; ***Endoh et al., 2008***). PRC2 proteins and the

H3K27me3 mark are proposed to maintain genes in a state in which they are poised for transcription and contribute to pluripotency (**Azuara et al., 2006**; **Boyer et al., 2006**; **Endoh et al., 2008**). In agreement with this notion, EZH2 is upregulated during reprogramming and EZH2 knock-down impairs reprogramming. Also, PRC2 is essential for pluripotency but is not essential in human ES cells in the 'naïve' state (**Shan et al., 2017**).

In ES cells, a significant proportion of polycomb target genes are repressed genes that encode transcription factors that are required for lineage specification later during development (**Boyer et al., 2006**; **Lee et al., 2006**). These genes are also co-occupied by the key pluripotency factors OCT4, NANOG, and SOX2, suggesting that the function of these pluripotency genes in repressing gene expression may be mediated by polycomb (**Boyer et al., 2005**; **Boyer et al., 2006**). Indeed, downregulation of the PRC1 and PRC2 activity in ES cells leads to global de-repression of these genes and to unscheduled differentiation (**Boyer et al., 2006**; **Endoh et al., 2008**). Deletion of one of the PRC2 components in mice (EED, SUZ12 or EZH2) results in severe defects in development, suggesting the mis-expression of lineage-specific genes (**Faust et al., 1995**; **O'Carroll et al., 2001**; **Pasini et al., 2004**). However, the deletion of PRC2 components in ES cells is not lethal (**Leeb et al., 2010**).

Interestingly, polycomb has also been reported to bind to bivalent genes, which are occupied by both the heterochromatic mark H3K27me3 and the active mark H3K4me3. In this scenario, PRC2 is important in maintaining lineage genes in a poised state ready to respond to differentiation cues (**Voigt et al., 2013**).

Telomeres are special heterochromatin structures at chromosome ends, which are formed by tandem repeats of the TTAGGG sequence bound by the so-called shelterin complex (**Blackburn, 2005**; **de Lange, 2005**; **Martínez and Blasco, 2011**). Telomeric chromatin is enriched in histone-repressive marks including H3K9me3 and H4K20me3 (**García-Cao et al., 2002**; **García-Cao et al., 2004**; **Gonzalo et al., 2005**; **Gonzalo et al., 2006**; **Benetti et al., 2007**; **Blasco, 2007**). Recently, we also reported that telomeres are enriched for the PRC2-repressive mark H3K27me3 (**Montero et al., 2018**). The function of telomeres is to protect chromosome ends from degradation and from triggering chromosomal aberrations, such as chromosome end-to-end fusions. Two independent studies showed that the shelterin component TRF1 is also greatly upregulated during reprogramming (**Boué et al., 2010**; **Schneider et al., 2013**). Indeed, we showed that TRF1 upregulation is an early event during cellular reprogramming, which precedes and is independent of telomere elongation by telomerase (**Schneider et al., 2013**). The *Terf1* gene is a direct target of OCT4, and is also essential for the induction and maintenance of pluripotency. In support of this, deletion of TRF1 causes embryonic lethality at the blastocyst stage (**Karlseder et al., 2003**). More recently, we showed that TRF1 is also upregulated during in vivo reprogramming, showing a similar pattern of expression to that of OCT4 in reprogrammed tissues (**Marión et al., 2017**). In spite of this solid evidence that TRF1 has an important role in pluripotency, the mechanisms that allow TRF1 to perform this mediating role have remained unknown until now.

PRC2 can interact both in vivo and in vitro with the long non-coding RNAs transcribed from telomeres, or TERRA, and this interaction is essential for the establishment of the H3K27me3 mark at telomeres (**Chu et al., 2017**; **Wang et al., 2017**; **Montero et al., 2018**). TERRA has also been shown to be associated with polycomb marks in the vicinity of genes and to modulate gene expression (**Chu et al., 2017**). Thus, there seems to be an interplay between telomere transcriptional status and long-range epigenetic regulation. In fact, PRC2 interacts with several long non-coding RNAs (lncRNAs), and this interaction is thought to regulate gene expression by recruiting PRC2 to specific loci. Some examples of lncRNAs that can physically interact with PRC2 and recruit it to specific loci include *Xist* (**Zhao et al., 2008**), *Hotair* (**Rinn et al., 2007**) and the antisense non-coding RNA in the *Cdkn2a* locus (**Yap et al., 2010**). These lncRNAs play important roles in X chromosome activation and tumorigenesis. However, how a lncRNA is able to provide specificity for PRC2 recruitment is not clear.

In addition, TERRA has been previously described to interact with the shelterin component TRF2, which can interact with TRF1, thus opening the possibility that polycomb may also be interacting with shelterin components. In this regard, a recent report showed that the *Arabidopsis thaliana* telomere-repeat binding factors (TRBs) recruit PRC proteins to different promoters through a telobox motif. In the absence of the three TRB proteins, the PRC2-mediated H3K27me3 mark was altered in

a similar manner to that of PRC2 mutants. Indeed, an interaction between TRB1–3 and PRC2 proteins was found (**Zhou et al., 2016b**; **Zhou et al., 2018**).

Here, we set to address the mechanisms through which OCT4-mediated TRF1 upregulation functions as an essential process for the induction and maintenance of pluripotency in mouse cells. To this end, we have used an unbiased genome-wide approach, looking for global changes in gene expression in the absence of TRF1. We make the unprecedented finding that TRF1 abrogation has a 'butterfly effect' on the transcription of naïve pluripotent cells, altering the epigenetic landscape of these cells through a novel mechanism, which involves TERRA-mediated polycomb recruitment to pluripotency genes and cell-fate genes.

Results

Abrogation of TRF1 in 2i-grown iPS cells changes the expression of genes related to pluripotency, differentiation and control by polycomb

To address whether TRF1 abrogation results in genome-wide changes in gene expression that could explain why TRF1 is required for pluripotency, we set to analyze the whole cellular transcriptome directly in induced pluripotent stem cells (iPS) cells in which TRF1 had been severely downregulated by the use of a short hairpin RNA (shRNA) (**Figure 1A**). We used *Trp53* (also known as p53)-null iPS cells to allow for cell proliferation in the absence of TRF1 (**Martínez et al., 2009**). We optimized the experiment so that the cells were harvested as soon as abrogation of TRF1 could be confirmed, in order to avoid deleterious effects associated to TRF1 depletion, such as increased DNA damage response. As expected from TRF1 downregulation, we observed an increase in the number of multitelomeric signals (MTS) (**Figure 1—figure supplement 1A**), which have previously been related to TRF1 abrogation (**Martínez et al., 2009**; **Sfeir et al., 2009**). In these conditions, we did not find a significant increase in the DNA damage marker γH2AX in TRF1-depleted cells compared to control cells using either western blot (**Figure 1—figure supplement 1B**) or immunofluorescence against γH2AX (**Figure 1—figure supplement 1C**). Similarly, we did not find any significant difference in DNA damage specifically located at telomeres, as determined by performing double immunofluorescence against γH2AX and the telomeric protein RAP1 together with quantification of the number of colocalizations or Telomere-Induced Foci (TIFs) (**Figure 1—figure supplement 1D**). These results indicate that the short period of cell harvesting after TRF1 depletion is not sufficient to allow the activation of a DNA damage response (DDR) that is greater than the basal levels shown by control cells, thus ruling out the possibility that the transcriptome effects observed here may be due to increased DNA damage resulting from TRF1 depletion. To determine the global gene expression changes that are associated with TRF1 depletion, we next performed RNA sequencing (RNA-seq) (see Materials and methods) in control *Trp53*-null 2i-grown iPS cells and *Trp53*-null 2i-grown iPS cells with downregulated TRF1 levels (**Figure 1A**). Gene set enrichment analysis (GSEA) revealed that genes that are downregulated when TRF1 is deleted, such as MYC, SOX2, NANOG and BMP (**Figure 1B–J**), are overrepresented in gene sets that are targets of pluripotency factors and pluripotency pathways. Interestingly, we found a clear enrichment in genes regulated by MYC (**Figure 1B**), concomitant with the fact that *Myc* was indeed downregulated by 3.8-fold in the absence of TRF1 in this RNA-seq experiment (Table S1). Also, genes that are downregulated when TRF1 is depleted were enriched among the gene targets of SOX2 and NANOG (**Figure 1C,D**), in genes that are upregulated in ES cells (**Figure 1E**) and in genes that are downregulated during the differentiation of embryoid bodies (**Figure 1F**). Together, these observations suggest that depletion of TRF1 induces the loss of pluripotency and initiates the differentiation of the iPS cells. In agreement with this notion, genes that are upregulated in the absence of TRF1 were enriched in gene sets that are upregulated during the differentiation of embryoid bodies from ES cells (**Figure 1H**), in targets of SUZ12 (**Figure 1I**), and in genes that are upregulated upon *knockout* of *Bmp2* (**Figure 1J**), a member of bone morphogenetic proteins that are important in tissue differentiation (**Zhou et al., 2016a**).

Further analysis of the RNA-seq results showed that 328 genes were downregulated and 483 genes upregulated significantly in the absence of TRF1 (**Figure 1K**). Analysis of the genes that are downregulated when abrogating TRF1 by Enrichr (a comprehensive gene set enrichment analysis web server) showed a significant enrichment in the targets of a large number of pluripotency

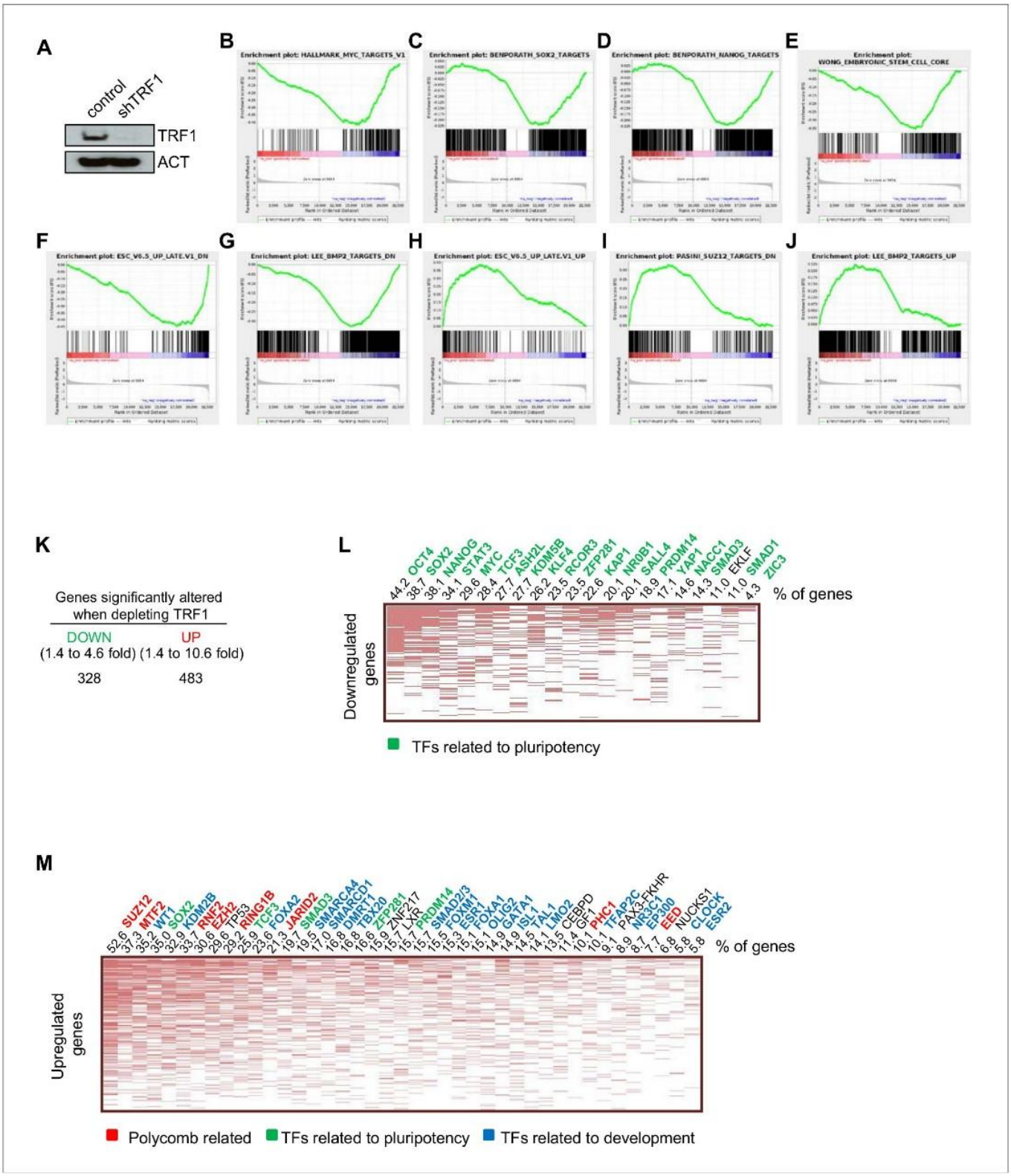


Figure 1. Transcriptome of TRF1-depleted iPS cells. (**A**) Western blot of TRF1 protein and actin (ACT) as the loading control in control and TRF1-depleted iPS cells. (**B–J**) Gene expression data obtained by RNA-seq of two independent experiments of TRF1-depletion in iPS cells, analyzed by Gene Set Enrichment Analysis (GSEA) to determine significantly enriched gene sets. Highly significant gene sets are shown here. Genes that were downregulated upon TRF1 deletion were enriched in targets of MYC (**B**), SOX2 (**C**), NANOG (**D**) and BMP2 (**G**). They were also enriched in

Figure 1 continued on next page

Figure 1 continued

genes that are expressed in ES cells (**E**) and in genes that are downregulated during differentiation (**F**). Genes that are upregulated upon TRF1 deletion were enriched in genes that are also upregulated during differentiation (**H**) and in targets of SUZ12 (**I**) and BMP2 (**J**). (**K**) Changes of gene expression in TRF1-depleted iPS cells analyzed by RNA-seq. The table summarizes the numbers of expressed transcripts that are expressed differentially in TRF1-depleted and control iPS cells. (**L**) Analysis using Enrichr of genes that are downregulated in TRF1-deleted iPS cells relative to control cells. The clustergram, representing the results of a CHEA analysis, shows that an important number of downregulated genes are targets of numerous pluripotency factors. The numbers on the top of the clustergram indicate the percentage of all of the downregulated genes that are targets of each pluripotency factor. Note that more than 40% of all downregulated genes are bound by OCT4. (**M**) Analysis using Enrichr of genes that are upregulated in TRF1-deleted iPS cells relative to control cells. The clustergram, representing the results of a CHEA analysis, shows that an important number of upregulated genes are mainly targets of polycomb-related proteins (red) and differentiation-related proteins (blue). The numbers on the top of the clustergram indicate the percentage of all of the upregulated genes that are targets of each factor.

DOI: https://doi.org/10.7554/eLife.44656.002

The following figure supplements are available for figure 1:

Figure supplement 1. TRF1 abrogation in 2i-grown iPS cells does not increase DNA damage.

DOI: https://doi.org/10.7554/eLife.44656.003

Figure supplement 2. Functional annotation of genes altered when depleting TRF1.

DOI: https://doi.org/10.7554/eLife.44656.004

factors, including NANOG, OCT4, KLF4, SOX2 and MYC, with *Myc* again being one of the most downregulated genes after depletion of TRF1 (***Figure 1L***). ***Figure 1L*** shows the percentages of the downregulated genes that are direct targets of the indicated transcription factors. Strikingly, up to 86% of the downregulated genes were targets of pluripotency factors. In particular, more than 40% of all downregulated genes were direct targets of OCT4 and more than 30% were direct targets of SOX2, NANOG and STAT3 (***Figure 1L***). These findings indicate that TRF1 depletion alters the expression of numerous genes that are controlled by key pluripotency transcription factors. In line with these findings, Enrichr analysis of the downregulated genes show that they were significantly enriched in genes encoding components of signaling pathways regulating the pluripotency of stem cells; 26 of these genes belong to PluriNetWork, which contains genes underlying pluripotency in mouse (***Figure 1—figure supplement 2A***). All together, these results indicate that TRF1 depletion in 2i-grown iPS cells reduces the expression of genes that are implicated in maintaining pluripotency programs.

Interestingly, similar Enrichr analysis of genes that are upregulated when depleting TRF1 (***Figure 1M***) showed a significant enrichment in the targets of the polycomb repressive complexes PRC1 and PRC2, together with pluripotency factors and factors related to development (***Figure 1M***). In particular, up to 82% were targets of differentiation factors, 64% were targets of PRC2, and 65% were targets of pluripotency factors. Wikipathways and KEGG pathways analysis revealed that the upregulated genes were also enriched in important pathways regulating pluripotency and differentiation, such as the BMP, Tgfβ and WNT signaling pathways (***Figure 1—figure supplement 2B*** left panel). In agreement with the unexpected finding that PRC2 targets were greatly enriched in TRF1-depleted cells, we found that the upregulated genes were enriched in the polycomb-related chromatin marks H3K27me3 and H3K9me3, which are characteristic of heterochromatin (***Figure 1—figure supplement 2B*** right panel). We also found that thesegenes are enriched in H3K4me3, a mark present in bivalent genes, where polycomb has been described to bind in primed ES cells (***Voigt et al., 2013***) (***Figure 1—figure supplement 2B*** right panel). These findings suggest that TRF1 depletion in 2i-grown iPS cell results in the re-activation of polycomb target genes and genes containing the polycomb repressive mark H3K27me3, a phenomenon associated with the loss of the pluripotency stage and the expression of lineage genes. Thus, pathways that are important in lineage commitment and pluripotency, such as the BMP, Tgfβ and WNT signaling pathways, are altered, consistent with loss of the pluripotency state. In summary, these findings indicate that abrogation of TRF1 expression in 2i-grown iPS cells alters the expression of genes that are related to pluripotency, differentiation, and polycomb complexes, supporting the hypothesis that the loss of TRF1 induces the loss of pluripotency and the induction of differentiation.

Abrogation of TRF1 in 2i-grown iPS cells alters global SUZ12 and H3K27me3 genomic distribution

As TRF1 deletion alters the expression of polycomb- and H3K27me3-regulated genes, we next set out to address whether TRF1 affects the global genome distribution of SUZ12 and H3K27me3 proteins. To do so, we performed chromatin immunoprecipitation of SUZ12 and H3K27me3 followed by deep sequencing (ChIP-seq) in control 2i-grown iPS cells and 2i-grown iPS cells in which TRF1 had been severely downregulated (***Figure 1A***). Interestingly, we found that TRF1 abrogation significantly increased the number of SUZ12 binding sites within the genome (4230 peaks) compared to that in control cells (440 peaks) (***Figure 2A***, and ***Figure 2—figure supplement 1***). Heat maps of the read distribution around the SUZ12 peaks confirmed a clear increase in SUZ12 deposition in these genomic regions upon TRF1 depletion, indicating that abrogation of TRF1 in 2i-grown iPS induces a strong recruitment of SUZ12 into the genome (***Figure 2B***). Total levels of SUZ12 protein were not altered upon TRF1 depletion in four independent experiments (***Figure 2C***, see bottom panel for quantification), indicating that the recruitment of SUZ12 to the genome is not due to overexpression of the SUZ12 protein. Then, we used Enrichr to analyze the characteristics of the genes that were associated with the genome sites where we observed an increased binding of SUZ12 protein upon depletion of TRF1 (see Clusters 1, 2 and 3 of the SUZ12 heatmap [***Figure 2B***]). Importantly, CHEA transcription factor analysis showed that the genes that were associated with sites of SUZ12 recruitment upon TRF1 deletion are the targets of a great number of pluripotency factors, including OCT4, NANOG, KLF4 and MYC (Table S2), as well as of components of the polycomb repressor complexes (Table S2). In addition, KEGG AND Wikipathway analysis showed that these genes are enriched in several pathways that are involved in the regulation of ESC pluripotency, such as the WNT and TGFβ pathways, pathways related to differentiation, and pathways related to the nervous system (***Figure 2D***, Table S3). Furthermore, analysis of GO Biological process indicated that these genes are highly involved in processes related to the nervous system (***Figure 2—figure supplement 2***). We performed a second ChIP-seq to confirm our results, and we found that the heatmaps of the two biological replicates of SUZ12 ChIP-seq, with reads plotted around the peaks for both replicates, show the same pattern (***Figure 2—figure supplement 1A***). Importantly, the analysis of the genomic regions in which the SUZ12 signal was increased was very similar (***Figure 2—figure supplement 1A***). Thus, we observed that all of the clusters of these heatmaps included regions where SUZ12 increases upon TRF1 depletion in both replicas (***Figure 2—figure supplement 1A***). Moreover, we showed that 99% of the genes that showed increased levels of SUZ12 upon TRF1 depletion in the first experiment also showed higher levels of SUZ12 in the second replicate, confirming the close similarity between both ChIP-seq replicates and the recruitment of SUZ12 to the described genomic sites (***Figure 2—figure supplement 1C***). These results suggest that TRF1 may control the expression of pluripotency- and differentiation-related proteins by directly or indirectly inducing the redistribution of the SUZ12 PRC2 component throughout the genome.

To further support our observation from ChIP-seq that TRF1 depletion increases PRC2 recruitment to the genome, we next performed an additional ChIP-seq analysis to study the H3K27me3 polycomb mark upon TRF1 abrogation. We found 14,317 H3K27me3 peaks in control 2i-grown iPS cells and 16,165 H3K27me3 peaks in TRF1-depleted 2i-grown iPS cells (***Figure 2E*** and ***Figure 2—figure supplement 1***). Heat maps of the read distribution around H3K27me3 peaks (***Figure 2F***) showed a cluster of peaks (Cluster 1) in which a clear increase of H3K27me3 deposition was observed. Total levels of H3K27me3 were not altered upon TRF1 depletion in four independent experiments (***Figure 2G***, see quantification in bottom panel), indicating that the deposition of H3K27me3 to those sites in the genome is not due to changes in the cellular levels of H3K27me3. We then use Enrichr to analyze the characteristics of the genes associated with the genome sites where we observed increased binding of H3K27me3 protein upon depletion of TRF1 (Cluster 1 of the H3K27me3 heatmap from ***Figure 2F***). KEGG and Wikipathway analyses showed results that were very similar to those of SUZ12, with a clear enrichment in pathways regulating pluripotency, differentiation, and the nervous system (***Figure 2H***). We performed a second ChIP-seq to confirm our results, and found that the heatmaps of the two biological replicates of H3K27me3 ChIP-seq, with reads plotted around the peaks for both replicates, show the same pattern (***Figure 2—figure supplement 1D***). We observed that clusters 3 and 7 of these heatmaps include regions where H3K27me3 increases upon TRF1 depletion in both replicates (***Figure 2—figure supplement 1D***).

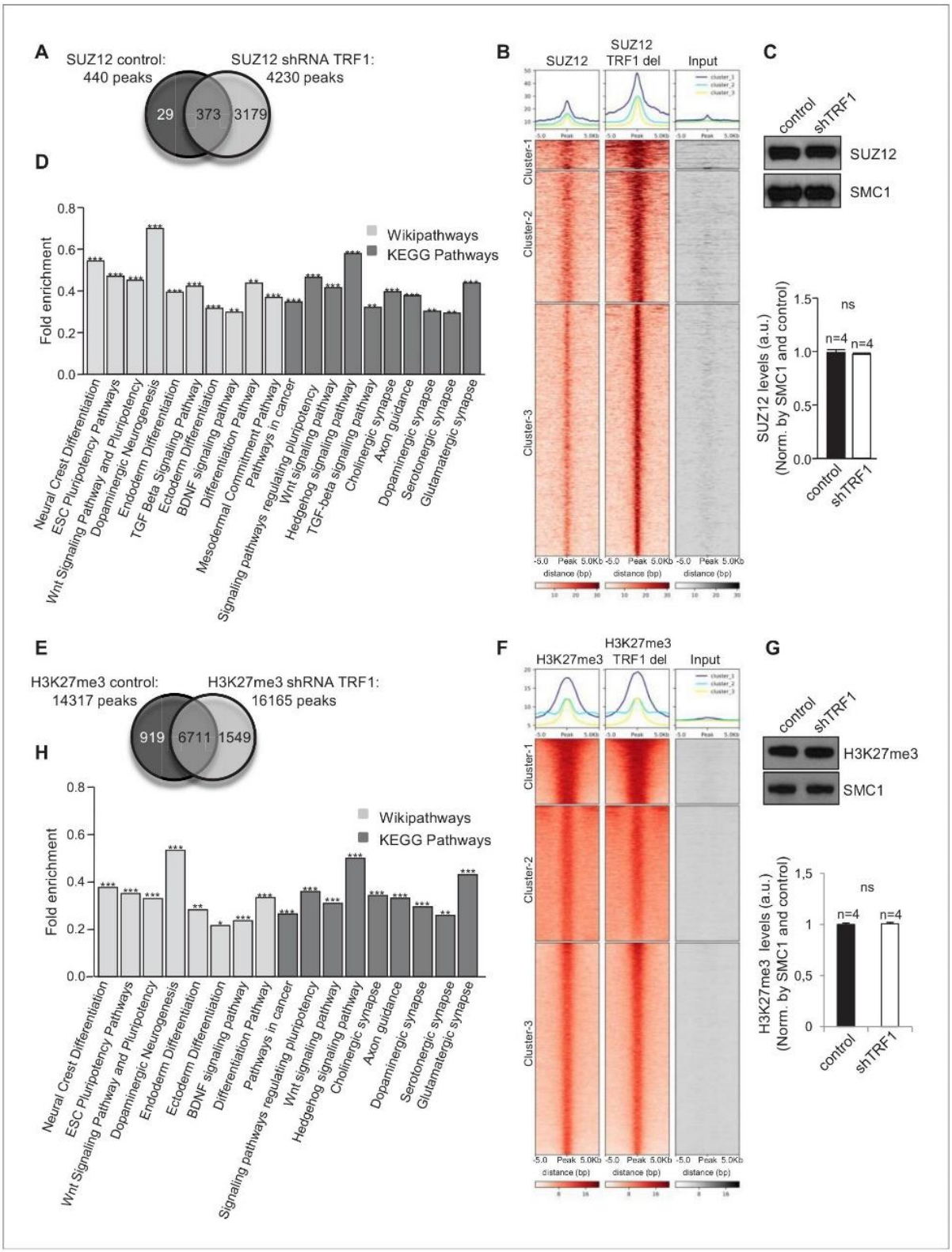


Figure 2. Abrogation of TRF1 alters SUZ12 and H3K27me3 genome localization. Analysis of the genome-wide binding of SUZ12 and H3K27me3 by chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) in control and TRF1-depleted 2i-grown iPS cells. (**A**) Number of SUZ12 binding peaks in control and TRF1-depleted cells. Note that TRF1 abrogation significantly increases the number of SUZ12 binding sites in the genome (4230 peaks) compared to that in control cells (440 peaks). The Venn diagram shows the number of genes that are annotated to SUZ12 peaks in both

Figure 2 continued on next page

Figure 2 continued

conditions. (**B**) Heat maps of the reads distribution around 5 Kb of SUZ12 peaks in control and TRF1-depleted cells. Note the visible increase in SUZ12 deposition in these sites upon TRF1 depletion. (**C**) (Top) Representative image of Western blots of SUZ12 and SMC1 (as loading control) from control and TRF1-depleted iPS cells. (Bottom) Quantification of SUZ12 levels in four independent experiments. Note that SUZ12 protein levels are not affected by the depletion of TRF1 protein. n = number of independent experiments. Error bars = SE. Statistical analysis, Student's t-test. (**D**) KEGG AND Wikipathways analysis of the genes associated with the sites of the genome were increased binding of SUZ12 protein was observed upon abrogation of TRF1. Genes annotated in Clusters 1, 2 and 3 of the SUZ12 heatmap (***Figure 2B***) were used. Note that these genes are mainly enriched in pathways related to pluripotency, differentiation and the development of the nervous system. ** = Adjusted p value <0.01, *** = Adjusted p value <0.001. (**E**) Number of H3K27me3 binding peaks in control and TRF1-deleted cells. Note that TRF1 abrogation increases the number of H3K27me3 binding sites within the genome (16,165 peaks) compared to that in control cells (14,317 peaks). The Venn diagram shows the number of genes annotated to H3K27me3 peaks in both conditions. (**F**) Heat maps of the reads distribution around 5 Kb of H3K27me3 peaks in control and TRF1-depleted cells. Note that Cluster 1 shows an increase in H3K27me3 deposition upon TRF1 downregulation. (**G**) (Top) Representative image of Western blots of H3K27me3 and SMC1 (as loading control) from control and TRF1-depleted iPS cells. (Bottom) Quantification of H3K27me3 levels in four independent experiments. n = number of independent experiments. Error bars = SE. Statistical analysis, Student's t-test. Note that H3K27me3 protein levels are not affected by the depletion of TRF1 protein. (**H**) KEGG and Wikipathways analysis of the genes associated with the genome sites were increased binding of H3K27me3 protein was observed upon depletion of TRF1. Genes annotated in Cluster 1 of the H3K27me3 heatmap (***Figure 2F***) were used. Note that the results are very similar to those obtained for SUZ12. ** = Adjusted p value <0.01, *** = Adjusted p value <0.001.

DOI: https://doi.org/10.7554/eLife.44656.005

The following figure supplements are available for figure 2:

Figure supplement 1. Validation of SUZ12 and H3K27me3 ChIP-seq.

DOI: https://doi.org/10.7554/eLife.44656.006

Figure supplement 2. Functional annotation analysis of genes enriched in SUZ12 upon TRF1 depletion.

DOI: https://doi.org/10.7554/eLife.44656.007

Figure supplement 3. Abrogation of TRF1 in 2i-grown iPS cells induces the loss of the naïve state and a transition to a primed or differentiated state.

DOI: https://doi.org/10.7554/eLife.44656.008

Figure supplement 4. Extratelomeric binding of TRF1 in 2i-grown iPS cells.

DOI: https://doi.org/10.7554/eLife.44656.009

Importantly, we showed that 90% of the genes that showed increased levels of H3K27me3 upon TRF1 depletion in the first experiment also showed higher levels of H3K27me3 in the second replicate, confirming the close similarity between the two ChIP-seq experiments (***Figure 2—figure supplement 1F***). All together, these results indicate that the depletion of TRF1 in 2i-grown iPS cells induces a dramatic recruitment of SUZ12 and H3K27me3 to genes controlling key pathways in pluripotency and differentiation.

Depletion of TRF1 induces the recruitment of SUZ12 and H3K27me3 to genes that are de-regulated in the absence of TRF1

To further understand how TRF1 influences pluripotency by affecting polycomb distribution, we made a Venn diagram showing genes that are bound by SUZ12 and H3K27me3 specifically in the absence of TRF1 and genes that are downregulated in the absence of TRF1 (***Figure 3A***). We found a set of 14 genes that were downregulated and that also recruited SUZ12 and H3K27me3 upon TRF1 abrogation (***Figure 3A***). Importantly, *Myc* was present within this set of genes, and showed a clear gain of SUZ12 and H3K27me3 peaks in the absence of TRF1 (***Figure 3B***). Of special interest in this set of genes were *Myc*, *Id1* and *Foxd3*, which are involved in control of pluripotency, the TGFβ pathway and the WNT pathway, respectively. These findings indicate that TRF1 abrogation induces the recruitment of SUZ12 and H3K27me3 to key pluripotency and development genes, thereby repressing their expression. We then selected a set of 30 genes that showed recruitment of either SUZ12 or H3K27me3, or of both SUZ12 and H3K27me3 when TRF1 was depleted and validated their expression by q-PCR in three TRF1-downregulation-independent experiments (***Figure 3C***). As control, we confirmed *Terf1* downregulation (***Figure 3C***). Importantly, we confirmed downregulation of all of the genes including *Myc*, thus validating the RNA-seq findings (***Figure 3C***). Next, to obtain a more accurate picture of SUZ12 and H3K27me3 binding to the genes that are downregulated in the absence of TRF1, we created heat maps of the distribution of SUZ12 and H3K27me3 reads within 2.5 Kb of the transcription start sites (TSS) of all of the genes that were significantly downregulated in the RNA-seq upon TRF1 abrogation (***Figure 3D***). We observed a cluster of genes with dramatic recruitment of SUZ12 to their TSS upon depletion of TRF1 (***Figure 3D***, Cluster 1 of SUZ12 heatmap),

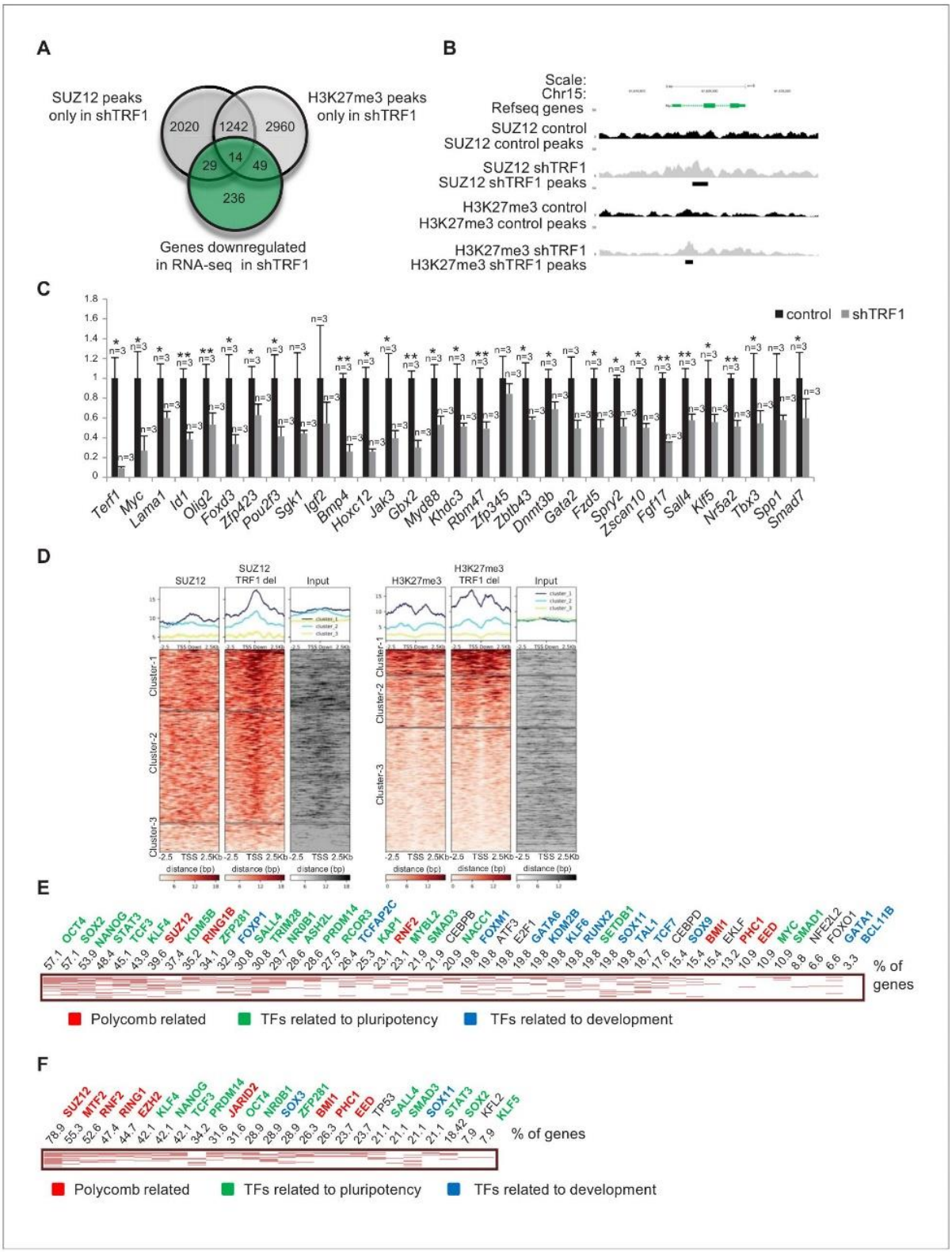


Figure 3. Depletion of TRF1 induces the recruitment of SUZ12 and H3K27me3 to genes that are downregulated in the absence of TRF1. (**A**) Venn diagram showing overlap of genes bound by SUZ12 and H3K27me3 specifically in the absence of TRF1 with the set of genes that are downregulated in this same condition, as obtained by RNA-seq. Note that a number of downregulated genes gain SUZ12 and H3K27me3 peaks upon TRF1 abrogation. (**B**) Gain of SUZ12 and H3K27me3 peaks in the *Myc* gene. (**C**) Gene expression analysis by q-PCR of genes that are downregulated upon TRF1

Figure 3 continued on next page

Figure 3 continued

depletion in which SUZ12 or H3K27me3 recruitment was detected. Note that abrogation of TRF1 induces a clear reduction in the expression of these genes, as observed before by RNA-seq. As a control, we confirmed that TRF1 expression was drastically decreased. n = number of independent experiments. Error bars = SE. Statistical analysis, one tail, paired Student's t-test. * = p value<0.05; ** = p value<0.01. (**D**) Heat maps of the distribution of reads of SUZ12 and H3K27me3 within 2.5 Kb of the transcription start sites (TSS) of all of the genes that are significantly downregulated in the RNA-seq upon TRF1 abrogation. Note that Clusters 1 of both the SUZ12 and the H3K27me3 heatmap show a clear recruitment of the corresponding protein. (**E**) Genes belonging to Cluster 1 of the Suz12 heatmap (***Figure 3D***), which present a dramatic enrichment of SUZ12 upon TRF1 depletion, were analyzed. The clustergram , representing the results of the CHEA analysis, shows that an important number of these genes are targets of numerous pluripotency factors (green). The numbers on the top of the clustergram indicate the percentage of all the downregulated genes that are targets of each factor. (**F**) Genes belonging to Cluster 1 of the H3K27me3 heatmap (***Figure 3D***), which present a clear enrichment of H3K27me3 upon TRF1 depletion, were analyzed. As in the case of SUZ12, the clustergram shows that an important number of these genes are targets of pluripotency factors (green).

DOI: https://doi.org/10.7554/eLife.44656.011

The following figure supplement is available for figure 3:

Figure supplement 1. Functional analysis of genes downregulated in the absence of TRF1 that show recruitment of SUZ12 and H3K27me3.

DOI: https://doi.org/10.7554/eLife.44656.012

and a cluster with a more moderate increase in SUZ12 binding (Cluster 2). For H3K27me3, we detected a cluster of genes (Cluster 1) that recruited H3K27me3 to their TSS upon depletion of TRF1 (***Figure 3D***). We then performed an Enrichr analysis of the genes included in Cluster 1 of the SUZ12 heatmaps, corresponding to those genes that present a clearer difference in binding of SUZ12 upon TRF1 depletion. The results showed that these genes were significantly enriched in direct targets of a high number of pluripotency factors, including OCT4, SOX2, NANOG and MYC, among others (***Figure 3E***). As expected, they were also enriched in direct targets of the polycomb complexes. Wikipathways and KEGG analysis showed that these genes were significantly enriched in important pathways controlling pluripotency and differentiation (***Figure 3—figure supplement 1A***). Analysis of the genes that are enriched in H3K27me3 upon depletion of TRF1 (Cluster 1 of H3K27me3 heatmap) showed similar results (***Figure 3F***, ***Figure 3—figure supplement 1B***). Altogether, these results indicate that depletion of TRF1 increases the binding of SUZ12 and the deposition of H3K27me3 at the TSS of genes that are downregulated when TRF1 is depleted, with a particular enrichment in pluripotency- and differentiation-related genes, thus reinforcing the notion that one of the mechanisms by which TRF1 influences pluripotency is through the polycomb repressive complex.

Similarly, we created a Venn diagram to overlap genes bound by SUZ12 and H3K27me3 specifically in the absence of TRF1 with the set of genes that are upregulated in the RNA-seq in this same condition (***Figure 4A***). Interestingly, we found a set of upregulated genes that were only bound by SUZ12 and H3K27me3 in the absence of TRF1. An example of such recruitment in gene *Cxcl12* is shown (***Figure 4B***). We confirmed the increased expression of some of these genes by q-PCR in three independent experiments of TRF1 abrogation (***Figure 4C***). As for the downregulated genes, to obtain more accurate information on SUZ12 and H3K27me3 binding to the genes that are upregulated in the absence of TRF1, we created heat maps of the reads distribution of SUZ12 and H3K27me3 within 2.5 Kb of the transcription start sites (TSS) of all of the genes that were significantly upregulated in the RNA-seq upon TRF1 abrogation (***Figure 4D***). Surprisingly, we observed recruitment of SUZ12 to a great number of upregulated genes (Clusters 1 and 2 of the SUZ12 heatmap (***Figure 4D***)). By contrast, recruitment of H3K27me3 was not so dramatic. We performed Enrichr analysis of the genes contained in Cluster 1 of the SUZ12 heatmap, which corresponds to the larger increase in SUZ12 recruitment. The results showed that upregulated genes contained in Cluster 1 are direct targets of differentiation factors, as well as pluripotency factors and components of polycomb complexes (***Figure 4E***). Importantly, these upregulated genes are enriched in pathways involved in cell differentiation and pluripotency (***Figure 4—figure supplement 1A***). Similar results were obtained when we analyzed genes in which H3K27me3 is recruited to the TSS (***Figure 4F***, ***Figure 4—figure supplement 1B***). These results indicate that depletion of TRF1 induces the recruitment of SUZ12 and the deposition of H3K27me3 to genes that are upregulated and that are involved in pluripotency and differentiation. These results are in agreement with previous work showing that SUZ12 is required for mouse embryonic stem cell differentiation and that

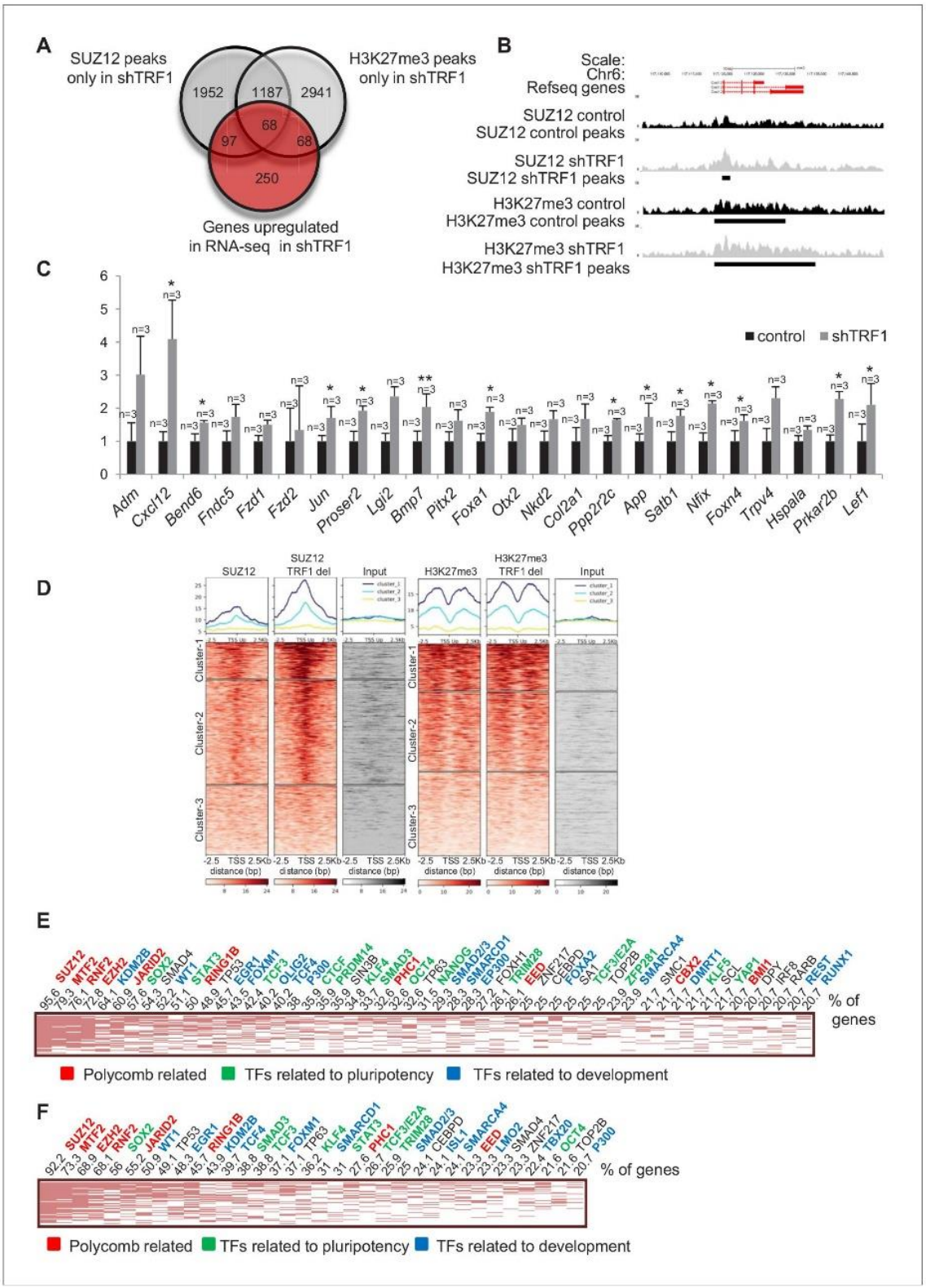


Figure 4. Depletion of TRF1 induces the recruitment of SUZ12 and H3K27me3 to genes that are upregulated in the absence of TRF1. (**A**) Venn diagram showing the overlap of genes bound by SUZ12 and H3K27me3 specifically in the absence of TRF1 with the set of genes that are upregulated in this same condition, as obtained by RNA-seq. Note that a number of upregulated genes gain SUZ12 and H3K27me3 peaks upon TRF1 abrogation. (**B**) Gain of SUZ12 and H3K27me3 peaks at the *Cxcl12* gene. (**C**) Gene expression analysis by q-PCR of genes that are upregulated upon TRF1 depletion in which

Figure 4 continued on next page

Figure 4 continued

SUZ12 or H3K27me3 recruitment was detected. Note that abrogation of TRF1 induces an increase in the expression of these genes, as observed previously by RNA-seq. n = number of independent experiments. Error bars = SE. Statistical analysis was carried out by one -tailed, paired Student's t-test. * = p value<0.05; ** = p value<0.01. (**D**) Heat maps of the reads distribution of SUZ12 and H3K27me3 within 2.5 Kb of the transcription start sites (TSS) of all of the genes that are significantly upregulated in the RNA-seq upon TRF1 abrogation. Note that a great number of upregulated genes (Clusters 1 and 2 of the SUZ12 heat map) show a clear increase in SUZ12 binding upon TRF1 abrogation, especially in Cluster 1 where the increase is dramatic. By contrast, only Cluster 1 of the H3K27me3 heatmap shows a moderate recruitment. Genes belonging to Cluster 1 of the SUZ12 heatmap (***Figure 4D***), which present a dramatic enrichment of SUZ12 upon TRF1 depletion, were analyzed. (**E**) The clustergram shows that these genes are targets of pluripotency factors (green), differentiation factors (blue) and encode components of the polycomb (red). Genes belonging to Cluster 1 of the H3K27me3 heatmap (***Figure 4D***), which present a moderate enrichment of H3K27me3 upon TRF1 depletion, were analyzed. (**F**) As in the case of SUZ12, the clustergram shows that an important number of these genes are targets of pluripotency factors (green), differentiation factors (blue) and components of the encode Polycomb (red).

DOI: https://doi.org/10.7554/eLife.44656.013

The following figure supplement is available for figure 4:

Figure supplement 1. Functional analysis of genes that are upregulated in the absence of TRF1 and that show recruitment of SUZ12 and H3K27me3.

DOI: https://doi.org/10.7554/eLife.44656.014

increased levels of PRC2 accumulate on a subset of genes during differentiation, despite their transcriptional activation (***Pasini et al., 2007***). It has been reported that SUZ12 binding to the genome of pluripotent cells in a naïve state is at a low level, which increases upon transferring these cells to serum (primed state) (***Marks et al., 2012***). Thus, we wondered whether the increased recruitment of SUZ12 to the genome upon depletion of TRF1 could represent the loss of the naïve state and a transition to a primed or differentiated state. In primed conditions, SUZ12 is bound to a set of so-called bivalent genes, including the *Hox* clusters (***Voigt et al., 2013***). Thus, we created heatmaps of the reads of SUZ12 and H3K27me3 within 2.5 Kb of the TSS of the described list of bivalent genes (***Marks et al., 2012***) (***Figure 2—figure supplement 3***). We found a dramatic increase of SUZ12 deposition at the TSS of bivalent genes, and a more moderate increase in H3K27me3 at these sites (***Figure 2—figure supplement 3A***). ***Figure 2—figure supplement 3B*** shows an example of this recruitment at some of the more relevant bivalent genes, the *Hox* clusters. We found that SUZ12 is absent from the *Hox* clusters when TRF1 is present, but is clearly recruited to them in the absence of TRF1. This finding is consistent with the idea that abrogation of TRF1 induces the loss of the naïve state. Also, analysis of the genes from Cluster 1 of the H3K27me3 heatmap in ***Figure 4D*** (genes that show a moderate increase in H3K27me3) shows that almost 70% of them belong to the group of bivalent genes. It has been reported that bivalent genes show a lower presence of H3K27me3 in naïve-state cells than in primed-state cells (***Marks et al., 2012***), once again supporting the idea that the depletion of TRF1 induces a transition from naïve to primed-differentiated state.

To further understand whether the contribution of TRF1 to polycomb localization was direct or indirect, we performed a chromatin immunoprecipitation of TRF1 followed by deep sequencing (ChIP-seq) in control 2i-grown iPS cells. We observed few peaks of TRF1 when comparing two replicates of the ChIP-seq experiment (***Figure 2—figure supplement 4A***). Comparison of the RNA-seq experiments with TRF1 ChIP-seq indicated that only three genes that are bound by TRF1 showed significantly altered expression when TRF1 was downregulated (***Figure 2—figure supplement 4A*** and ***Figure 2—figure supplement 4C***). These genes were *Lama1*, which was significantly downregulated when TRF1 was deleted in our RNA sequencing experiment (***Figure 3C***), and *Pitx2* and *Fam43a*, which were significantly upregulated when TRF1 was depleted (***Figure 4C*** and Table S1). Interestingly, the majority of the TRF1 peaks outside the telomere were located at extra-telomeric repetitions (TTAGGG or CCCTAA) (***Figure 2—figure supplement 4B***). Enrichr analysis of the genes that are associated with TRF1 peaks revealed that four of these genes (***Figure 2—figure supplement 4B***, labeled in bold), were targets of ZFP322A, a protein that is expressed in the *Inner Cell Mass*, which is essential for ES cell pluripotency and that binds to *Oct4* and *Nanog* promoters and regulates their transcription (***Ma et al., 2014***). These four genes were present among the 100 binding sites with top-ranked peak heights in ZFP322A ChIP-seq analysis (***Ma et al., 2014***). Most of the TRF1 peaks are located in intergenic regions, but in all four of the genes targetted by ZFP322A, the TRF1 peaks were located in the introns. Two of these genes (*Pde1c* and *Ppp1r9a*) were upregulated in the RNA-seq and one of them, *Lama1*, was significantly downregulated (***Figure 2—figure***

supplement 4C), whereas expression of the fourth gene (*Bbox1*) was not detected. Interestingly, ZFP322A controls the expression of *Sall4* and *Zscan10*, two genes that are downregulated upon TRF1 deletion (see *Figure 3C*). These results suggest that TRF1 may directly regulate ZFP322A, which in turn, regulates key genes involved in pluripotency, although the mRNA levels of this gene were not altered by TRF1 abrogation in the ChIP-seq data (not shown). However, although TRF1 is able to bind a few locations in the genome, this binding does not seem sufficient to explain the vast epigenetic changes that we observed in the absence of TRF1.

In order to determine whether TRF1 directly regulates the binding of polycomb in the genome, we performed mass spectroscopy of protein complexes that were immunoprecipitated with an anti-GFP antibody in $Terf1^{GFP/GFP}$, $Terf1^{+/GFP}$ and $Terf1^{+/+}$ iPS cells (*Schneider et al., 2013*). We identified 64 proteins that showed co-immunoprecipitation in both $Terf1^{GFP/GFP}$ and $Terf1^{+/GFP}$ cells but not in $Terf1^{+/+}$ cells. We confirmed the presence in this group of proteins of different components of shelterin, namely TPP1, RAP1, TRF1, TIN2 and POT1. We did not, however, detect the presence of SUZ12 in the immunocomplex, suggesting that there is no direct interaction between TRF1 and PRC2, at least in this experimental setting (*Supplementary file 4*).

Depletion of TRF1 in 2i-grown iPS cells induces the upregulation of TERRA RNAs expression

Interestingly, the telomeric TERRA RNAs have been recently shown to bind throughout the genome in H3K27me3-rich regions, and to interact with components of the polycomb complex (*Chu et al., 2017*). In our group, we also recently showed that TERRA proteins are required to recruit polycomb to telomeres and so contribute to H3K27me3 deposition at telomeric chromatin (*Montero et al., 2018*). With these findings in mind, we set to address whether the expression of TERRA was altered when TRF1 was depleted, which could explain, at least in part, the effects of TRF1 depletion on polycomb. First, we performed RNA-FISH analysis using a TERRA-specific probe in control 2i-grown iPS cells and in 2i-grown iPS cells in which TRF1 expression had been reduced. The results showed a dramatic upregulation of TERRA signal in TRF1-depleted cells (*Figure 5A*). Northern-blot analysis of TERRA levels in three independent TRF1 abrogation experiments in iPS cells showed TERRA upregulation as a consequence of TRF1 downregulation (*Figure 5B*). Note that TERRA is normally increased in iPS cells compared to differentiated MEFs cells, as we have previously described (*Marion et al., 2009*), and that TERRA levels are further increased by TRF1 abrogation in iPS cells (*Figure 5C*), clearly indicating that TRF1 has a role as a repressor of TERRA during the induction of pluripotency. Finally, we also confirmed increased TERRA expression as the result of TRF1 abrogation in keratinocytes from $Terf1^{\Delta/\Delta}$ *K5-Cre* newborn mice, which lack TRF1 expression (*Figure 5D*). These results are in agreement with previous findings from other groups in human cells (*Porro et al., 2014*; *Zeng et al., 2017*; *Sadhukhan et al., 2018*), which also show TERRA increase upon TRF1 depletion. Other studies, however, showed that depletion of TRF1 in mouse immortalized MEFs did not increase TERRA levels (*Sfeir et al., 2009*), and that downregulation of TRF1 by siRNA in the immortalized mouse myoblast cell line C2C12 reduced the expression of TERRA (*Schoeftner and Blasco, 2008*). A possible explanation for this apparent contradiction could be that the immortalization process may alter TERRA expression or that TERRA regulation by TRF1 is cell-type specific. In any case, the results shown here clearly demonstrate a dramatic increase in TERRA in two different mouse cells types (iPS cells and keratynocytes) upon TRF1 depletion (*Figure 5*), which in turn could contribute to the recruitment of polycomb to new parts of the genome.

Previous studies have demonstrated that the depletion of the shelterin component TRF2 can also induce higher TERRA expression (*Porro et al., 2014*; *Rossiello et al., 2017*). We wondered whether the global changes in gene expression observed upon TRF1 abrogation in 2i-grown iPS cells were also detected when depleting TRF2 or whether they were specific for TRF1 depletion. To this end, we downregulated the expression of TRF2 by means of an shRNA (*Figure 5—figure supplement 1A*) in $Trp53^{-/-}$2i-grown iPS cells and measured TERRA levels by RNA-FISH analysis, using a TERRA-specific probe (*Figure 5—figure supplement 1B*), and by Northern Blot (*Figure 5—figure supplement 1C*). However, in both cases we found that depletion of TRF2 in our experimental system did not induce an increase in TERRA levels (*Figure 5—figure supplement 1B–C*), supporting the idea that, at least in 2i-grown iPS cells, TRF2 does not have a major role in controlling TERRA expression.

We have previously shown that TERRA is upregulated during the induction of iPS cells (*Marion et al., 2009*) (see also *Figure 5*). We wondered whether TERRA levels were also regulated

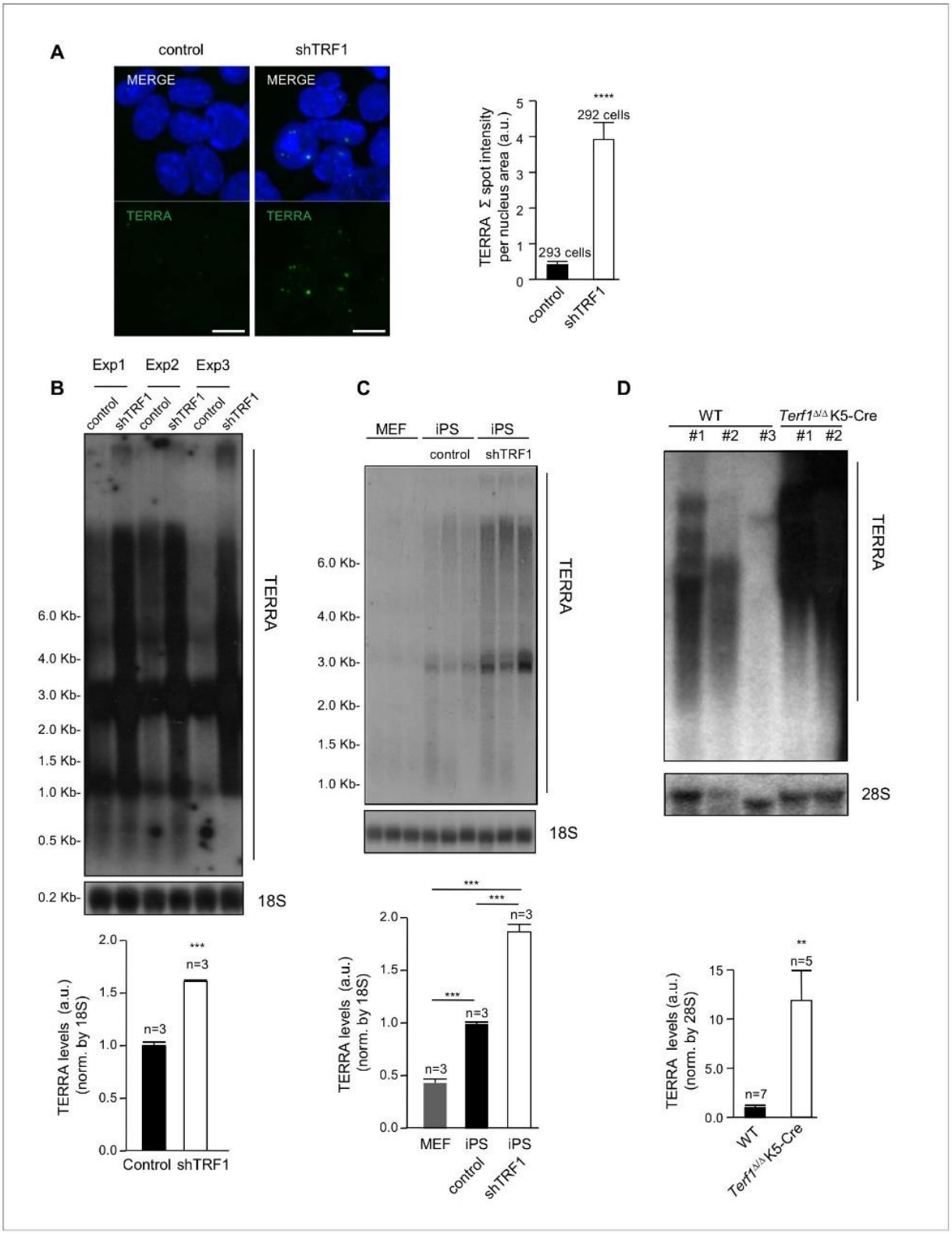


Figure 5. Abrogation of TRF1 induces the upregulation of TERRA RNAs expression. (**A**) Left, representative image of RNA-FISH using a TERRA-specific probe in control 2i-grown iPS and in 2i-grown iPS cells in which TRF1 is depleted. Note the dramatic increase in TERRA signal in cells lacking TRF1. Right, quantification of the RNA-FISH images. Data were obtained from one experiment. n = number of cells analyzed from each sample. Error bars = SE. Statistical analysis was carried out by Student's t-test. **** = p-value <0.0001. (**B**) Top, representative image of a Northern-blot analysis of

Figure 5 continued on next page

Figure 5 continued

TERRA RNAs expression in control 2i-grown iPS and in 2i-grown iPS cells in which TRF1 is depleted in three independent experiments. Note the clear upregulation of TERRA levels in cells lacking TRF1, confirming the RNA-FISH data. Bottom, quantification of TERRA levels from the Northern blot, normalized to 18S levels. Error bars = SE. Statistical analysis was carried out by Student's t-test. *** = p-value <0.0001. (**C**) Top, image of a Northern-blot analysis of TERRA RNAs expression in MEF, 2i-grown iPS and 2i-grown iPS depleted for TRF1. Bottom, quantification of TERRA levels from the Northern blot, normalized to 18S levels. n = number of independent experiments. Error bars = SE. Statistical analysis was carried out by Student's t-test. (**D**) Top, representative image of a Northern-blot analysis of TERRA RNAs expression in keratinocytes of wild-type or *Terf1*$^{\Delta/\Delta}$ *K5-Cre* (lacking expression of TRF1) newborn mice. Again, note the clear upregulation of TERRA levels in cells lacking TRF1. Bottom, quantification of TERRA levels from the Northern blot, normalized to 28S levels. n = number of independent newborn mice analyzed. Error bars = SE. Statistical analysis was carried out by Student's t-test.

DOI: https://doi.org/10.7554/eLife.44656.015

The following figure supplements are available for figure 5:

Figure supplement 1. TRF2 abrogation in *Trp53*$^{-/-}$ 2i-grown iPS cells does not change TERRA levels.

DOI: https://doi.org/10.7554/eLife.44656.016

Figure supplement 2. TRF1 protein and TERRA levels decrease upon differentiation of ES cells.

DOI: https://doi.org/10.7554/eLife.44656.017

during the process of differentiation. To answer this question, we differentiated ES cells with retinoic acid and measured TERRA levels by RNA-FISH after 5 days of treatment (***Figure 5—figure supplement 2A***). In agreement with loss of pluripotency, we confirmed significantly decreased *Nanog* levels after 5 days of treatment with retinoic acid (***Figure 5—figure supplement 2A***). We found that TERRA expression decreased during cellular differentiation (***Figure 5—figure supplement 2B***). We also found that upon differentiation of ES cells with retinoic acid, the levels of TRF1 protein were decreased, as measured by immunofluorescence (***Figure 5—figure supplement 2C***), in accordance with previous data showing higher levels of TRF1 in pluripotent cells (***Marion et al., 2009***; ***Schneider et al., 2013***). These findings suggest a fine-tuned regulation of TERRA levels during development, which is most probably linked to cell identity.

Genomic TERRA binding correlates with genes that are differentially expressed in the absence of TRF1 and at which SUZ12 and H3K27me3 are present

In order to address a role of TERRA in PRC2 localization to genes, we analyzed TERRA binding throughout the whole genome in our experimental setting. To this end, we performed TERRA CHIRT sequencing (CHIRT-seq) (as described by ***Chu et al., 2017***), in 2i-grown *Trp53*-null iPS cells. We found that TERRA was able to bind to 10,670 locations across the genome with an enrichment fold equal to or higher than 10 when compared to the input. Using Homer software, TERRA CHIRT peaks with an enrichment fold of 10 or higher were assigned to genes and genome regions. We found that the vast majority of TERRA binding sites were localized in non-coding regions (mostly in intergenic or intronic regions) (***Figure 6A***), in agreement with previous findings in embryonic stem cells (ESC) (***Chu et al., 2017***). Importantly, we found that 90.4% of the genes previously associated to TERRA peaks in ESC (***Chu et al., 2017***) were also found in our CHIRT-seq with 2i-grown iPS cells (***Figure 6B***), thus validating both TERRA binding sites along the genome and our CHIRT experiment. Importantly, we found that the TERRA peaks with a higher fold enrichment were localized at the subtelomere of chromosome 18 (***Figure 6C***), a locus previously described by us as one of the *bona fide* TERRA loci in murine cells (***López de Silanes et al., 2014***), thus further validating the importance of this locus in TERRA transcription. Also, the peaks with a higher fold enrichment colocalized with the subtelomeric sequence of TERRA at chromosome 18 (***Figure 6C***), confirming that we can capture the cis interaction of TERRA with the chromatin by using CHIRT. Altogether, these data clearly show that the CHIRT-seq technique was able to map TERRA localizations in the genome, including those at chromosome 18 subtelomere that were previously described by us as one of the main origins of mouse TERRA.

We next analyzed the genes that are associated with TERRA peaks using Enrichr. We found that more than 40% of these genes were targets of SUZ12, and that they were also bound by several components of the PRC2 complex (***Figure 6D***), supporting the idea that TERRA and PRC2 bound to similar locations in the genome. Furthermore, most of those genes were bound by transcription

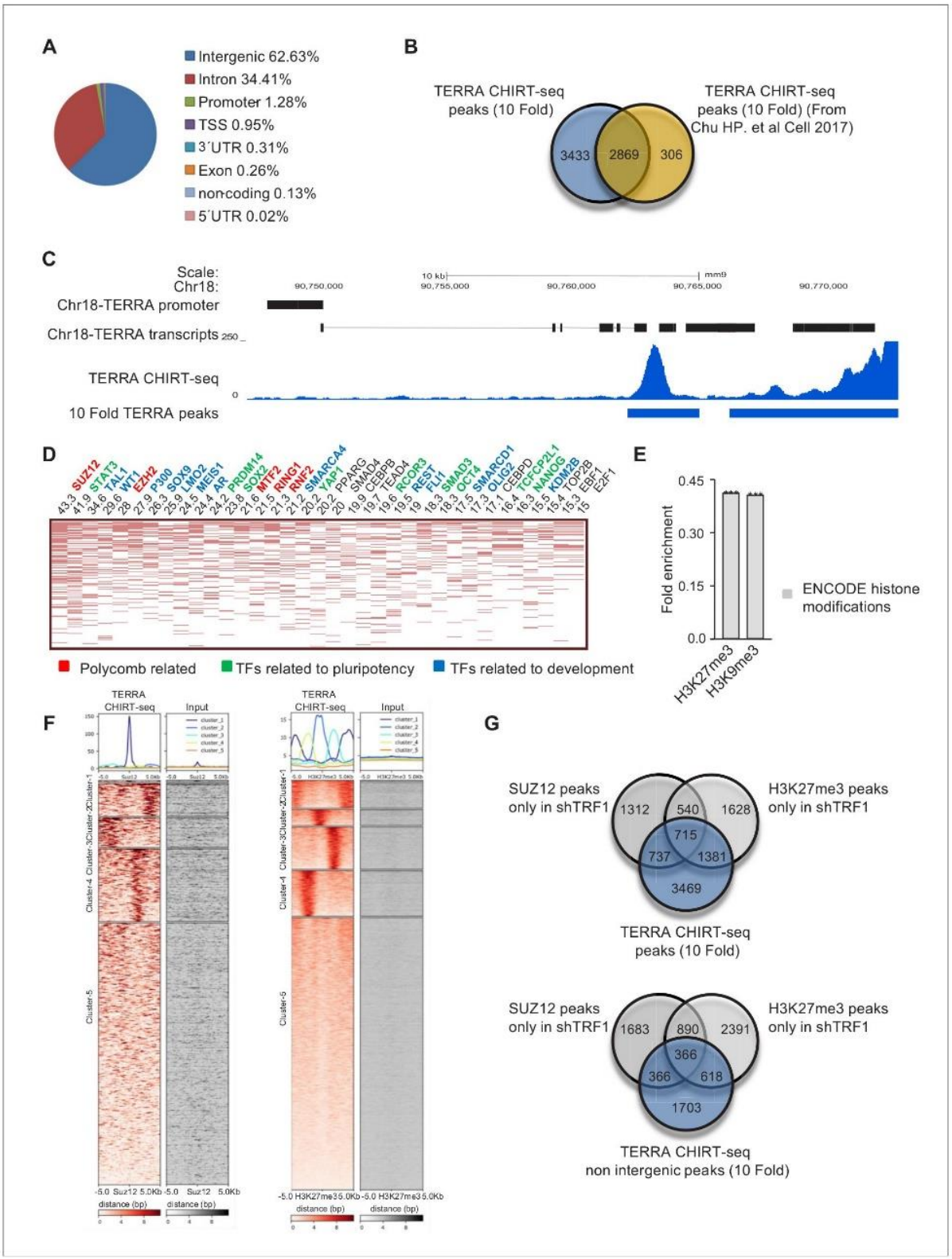


Figure 6. TERRA CHIRT-seq analysis. (**A**) Genome-wide TERRA localization in different genome regions. (**B**) Comparison of the genes associated with TERRA peaks in this work and in previously described work (***Chu et al., 2017***). Note that the majority of the genes found in the previous study are present in our CHIRT-seq. (**C**) Representation of the peak with higher enrichment in TERRA CHIRT-seq. Note that this peak coincides with a TERRA locus previously described at chromosome 18 (***López de Silanes et al., 2014***). (**D**) CHEA analysis of genes annotated to TERRA peaks. (**E**) Encode

Figure 6 continued on next page

Figure 6 continued

histone modifications analysis of genes annotated to TERRA peaks. (**F**) Heatmaps of TERRA reads within 5 Kb of SUZ12 and H3K27me3 peaks. (**G**) Top, overlapping of genes annotated to TERRA peaks with the peaks of SUZ12 and H3K27me3 exclusive for the shTRF1 ChIP-seq sample. Bottom, overlapping of genes annotated to TERRA peaks in non intergenic regions, with peaks of SUZ12 and H3K27me3 exclusive for the shTRF1 ChIP-seq sample. The experiment was performed once.

DOI: https://doi.org/10.7554/eLife.44656.018

factors that are related to pluripotency or differentiation (***Figure 6D***), suggesting that TERRA could be important for the loss of pluripotency after TRF1 depletion, through PRC2 regulation. In agreement with this, those genes were enriched in the H3K27me3 mark (***Figure 6E***).

We then performed heatmaps of the TERRA CHIRT-seq reads within 5 Kb of the peaks from SUZ12 and H3K27me3 ChIP-seq (***Figure 6F***). We found that 34.4% of the SUZ12 peaks and 32.6% of the H3K27me3 peaks have TERRA binding in their vicinity (***Figure 6F***). To understand whether the gain of SUZ12 and H3K27me3 binding after TRF1 removal is associated to locations where TERRA is present, we compared the genes that are associated with TERRA peaks with the genes associated with SUZ12 and H3K27me3 peaks that are exclusive for the shTRF1 ChIP-seq (***Figure 6G***). We found that 43.9% of the genes annotated to SUZ12 peaks were exclusive for shTRF1 ChIP-seq and that 49.2% of those from H3K27me3 peaks were also annotated in the TERRA CHIRT-seq. The fact that only a fraction of the newly recruited sites for SUZ12 coincides with TERRA peaks could be explained if SUZ12 recruitment to the genome were to be induced in part by the secondary changes in global epigenetic status and gene expression generated by TRF1 depletion. Also, the chromatin sonication process required for the ChIP and CHIRT techniques may be disrupting the long-range interaction mediated by TERRA. All together, these data indicate that there is a correlation of TERRA location and SUZ12 and H3K27me3 that includes the genes in which SUZ12 and H3K27me3 are deposited after TRF1 depletion, suggesting that TERRA could be mediating the effects of TRF1 abrogation on polycomb-regulated genes.

To evaluate whether the presence of TERRA binding could be associated with the changes of gene expression observed after TRF1 depletion, we compared the genes that were significantly altered in our RNA-seq with the genes that are associated with TERRA peaks (***Figure 7A***). We found that 29.9% of the genes that are downregulated and 33.5% of the genes that are upregulated when TRF1 was deleted are annotated to TERRA peaks. Then, we created heatmaps of the TERRA CHIRT-seq reads within 2.5 Kb of the TSS of the downregulated genes (***Figure 7B***) and the upregulated genes (***Figure 7C***). We observed that 29.1% of the downregulated genes (Clusters 1 and 2 ***Figure 7B***) and 27% of the upregulated genes (Clusters 1 and 2 ***Figure 7C***) are bound by TERRA in the proximity of the TSS. To determine whether those genes have increased binding of SUZ12 or an increased number of H3K27me3 marks after TRF1 removal, we compared the downregulated and upregulated genes that present TERRA signal (***Figure 7B–C***) with the genes with increased SUZ12 or H3K27me3 signal (from ***Figure 3D*** and ***Figure 4D***). The results confirmed that 90.1% of the downregulated genes bound by TERRA and 70.1% of the upregulated ones have increased SUZ12 or H3K27me3 marks (as determined by the reads of the heatmaps) (***Figure 7D,E***). As shown in the UCSC genome browser representation of the SUZ12 and H3K27me3 CHIP-seq and the TERRA CHIRT-seq of the downregulated gene *Pou2f3* and of the upregulated gene *Bmp7* (***Figure 7D,E***), the TERRA peaks are present at the promoters of the genes and close to the increased SUZ12 or H3K27me3 signal. These data suggest that a significant number of the transcriptional changes observed after TRF1 deletion could be due to the recruitment of PRC2 complex by TERRA.

Next, we analyzed the genes that are co-regulated by the PRC2 complex and TERRA using Enrichr. We found that genes that were downregulated when TRF1 was depleted and that had increased binding of PRC2 and TERRA (***Figure 7D***) are mostly targets of important transcription factors involved in pluripotency, such as OCT4, SOX2 or NANOG (***Figure 7F***). Also, they are significantly enriched in the PluriNetwork pathway from the Wikipathways database (***Figure 7G***). In the case of genes that are upregulated when TRF1 was depleted and that had increased binding of PRC2 and TERRA (***Figure 7E***), the Enrichr analysis revealed that they were targets of several differentiation transcription factors, SUZ12 and other PRC2 complex components, and to a less extent, of transcription factors involved in pluripotency (***Figure 7H***). Moreover, those genes are enriched in the Neural Crest Differentiation pathway from Wikipathways (***Figure 7I***) and in genes in which the

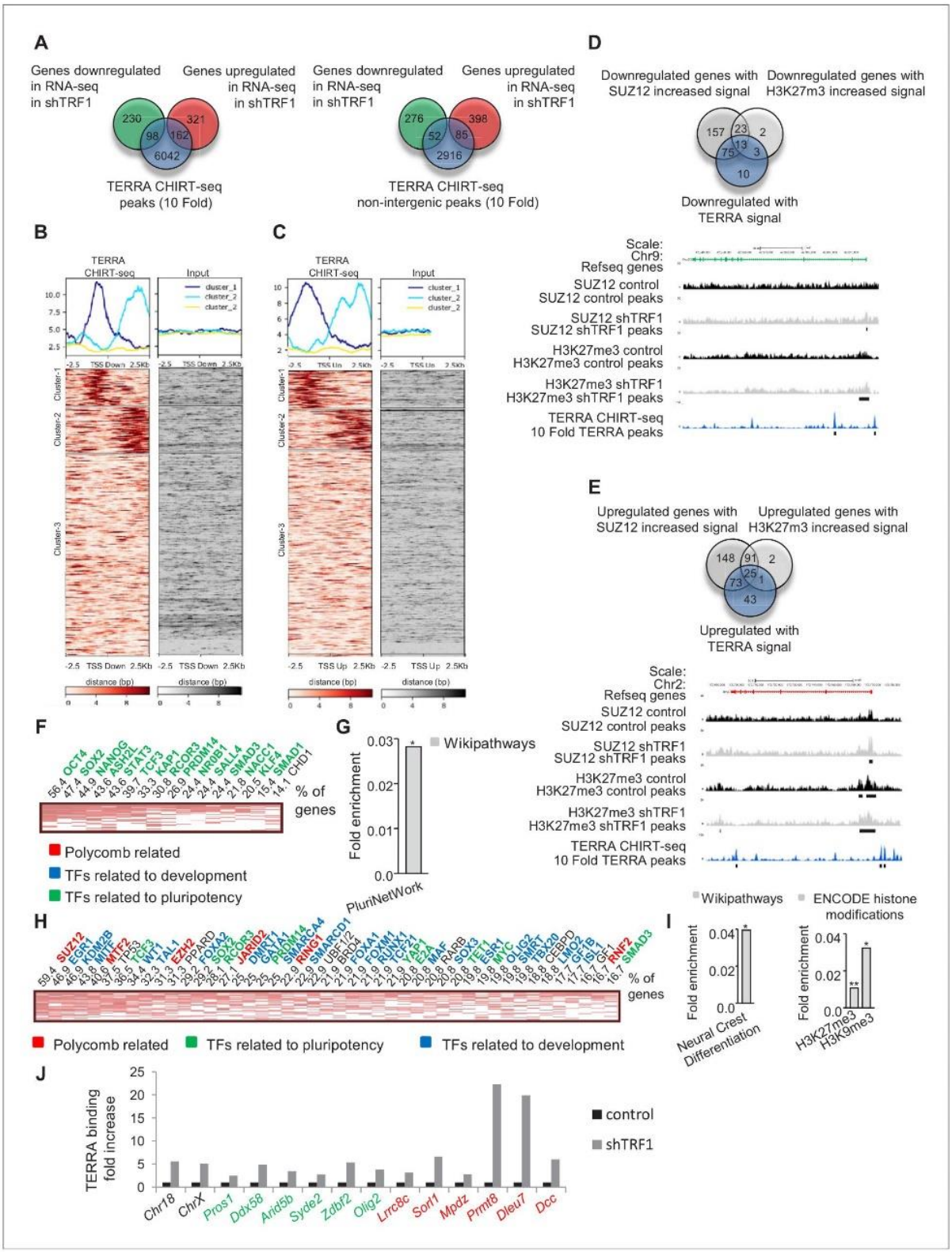


Figure 7. TERRA binding correlates with genes that are differentially expressed in the absence of TRF1 and with the presence of SUZ12 and H3K27me3 and is increased upon TRF1 depletion. (**A**) Left, Venn diagram showing the overlap between the genes annotated to TERRA peaks and genes that are up- or downregulated in the RNA-seq after TRF1 depletion. Right, Venn diagram showing the overlap between the genes annotated to TERRA peaks in non-intergenic regions and genes that are up- or downregulated in the RNA-seq after TRF1 depletion. (**B**) Heatmaps showing TERRA CHIRT-

Figure 7 continued on next page

Figure 7 continued

seq reads within 2.5 Kb of TSS of genes downregulated after TRF1 depletion. (**C**) Heatmaps showing TERRA CHIRT-seq reads within 2.5 Kb of TSS of genes upregulated after TRF1 depletion. (**D**) (Top) Overlapping between downregulated genes with TERRA signal and downregulated genes in which the SUZ12 or H3K27me3 signal is increased. (Bottom) Representative image of SUZ12, H3K27me3 and TERRA reads and peaks in a downregulated gene. (**E**) (Top) Overlapping between upregulated genes with TERRA signal and upregulated genes in which the SUZ12 or H3K27me3 signal is increased. (Bottom) Representative image of SUZ12, H3K27me3 and TERRA reads and peaks in an upregulated gene. (**F**) CHEA analysis of genes downregulated in the RNA-seq after the deletion of TRF1 that are bound by TERRA and that show increased SUZ12 and/or H3K27me3 signal. (**G**) Wikipathways analysis of genes from panel (F). (**H**) CHEA analysis of genes upregulated in the RNA-seq after deletion of TRF1 that are bound by TERRA and that show increased SUZ12 and/or H3K27me3 signal. (**I**) Wikipathways analysis and Encode histone modification analysis of genes from panel (H). (**J**) CHIRT-qPCR in control and shTRF1 samples from one experiment in genomic regions previously identified as TERRA binding sites in the CHIRT-seq. Genomic regions labeled in black correspond to subtelomeric regions where TERRA is potentially transcribed. Genomic regions labeled in green or red correspond to regions close to the promoters of genes that are downregulated or upregulated, respectively, when downregulating TRF1. Note that, in all the cases, TERRA binding is clearly increased upon TRF1 abrogation. Values were normalized to inputs and to a genomic region where TERRA binding was not found in the CHIRT-seq.

DOI: https://doi.org/10.7554/eLife.44656.019

ENCODE histone modifications database shows the presence of H3K27me3 and H3K9m3 marks (***Figure 7I***). These data indicate that the possible recruitment of PRC2 by TERRA after TRF1 depletion could alter a set of genes that can initiate the process of naïve state loss and differentiation.

TRF1 depletion induces increased TERRA binding to genes with increased polycomb binding and altered gene expression upon TRF1 abrogation

To test whether TRF1 abrogation resulted in increased recruitment of PRC2 by TERRA, we selected a number of genomic regions close to gene promoters where we had previously found both binding of TERRA in our CHIRT-seq and increased binding of PRC2 and changes in gene expression upon TRF1 depletion. We performed CHIRT-qPCR analysis for these regions in both 2i-grown iPS depleted for TRF1 and in non-depleted controls (***Figure 7J***). As a control, we included two genomic regions where TERRA is potentially transcribed (***Figure 7J***, labeled in black). We found a clear increase in TERRA binding to all of these regions upon TRF1 depletion, including the control regions where TERRA is potentially transcribed (suggesting that the increased levels of TERRA upon TRF1 deletion result from higher levels of TERRA transcription), the downregulated genes (***Figure 7J***, labeled in green) and the upregulated genes (***Figure 7J***, labeled in red). These results indicate both higher TERRA binding to genes with increased binding of PRC2 and changes in gene expression upon TRF1 depletion, suggesting that the higher TERRA binding could mediate the increased SUZ12 recruitment in these regions, and could thus control the expression of genes that are important for pluripotency and differentiation (see Model in (***Figure 8***).

As higher levels of TERRA mediate the increased recruitment of SUZ12 to the genome, we wondered whether TRF1 depletion would increase the interaction between TERRA and SUZ12. We therefore performed an RNA immunoprecipitation (RIP) experiment, in which we immunoprecipitated SUZ12 in both control and TRF1-depleted *Trp53*$^{-/-}$ 2i-grown iPS cells, and measured the amount of TERRA that was pulled-down with SUZ12. The results indicate that, indeed, SUZ12 shows a higher level of binding to TERRA when TRF1 is depleted (***Figure 9***), reinforcing the idea that higher levels of TERRA mediate the recruitment of SUZ12 to the genome.

Discussion

In recent years, mounting evidence has suggested a role of the TRF1 telomere binding protein in the acquisition and maintenance of pluripotency and stemness, which seemed to be additional to its known role in maintaining telomere protection and preventing telomere fusions. In particular, we and others found that TRF1 is highly upregulated in ES cells and iPS cells (***Boué et al., 2010***; ***Schneider et al., 2013***). Our group further showed that *Terf1* is a direct target of the pluripotency gene *OCT4* and that the upregulation of *Terf1* is an early event during the induction of pluripotency (***Schneider et al., 2013***). Indeed, we found that TRF1 is needed for both the induction and the maintenance of pluripotency, being essential therefore for nuclear reprogramming

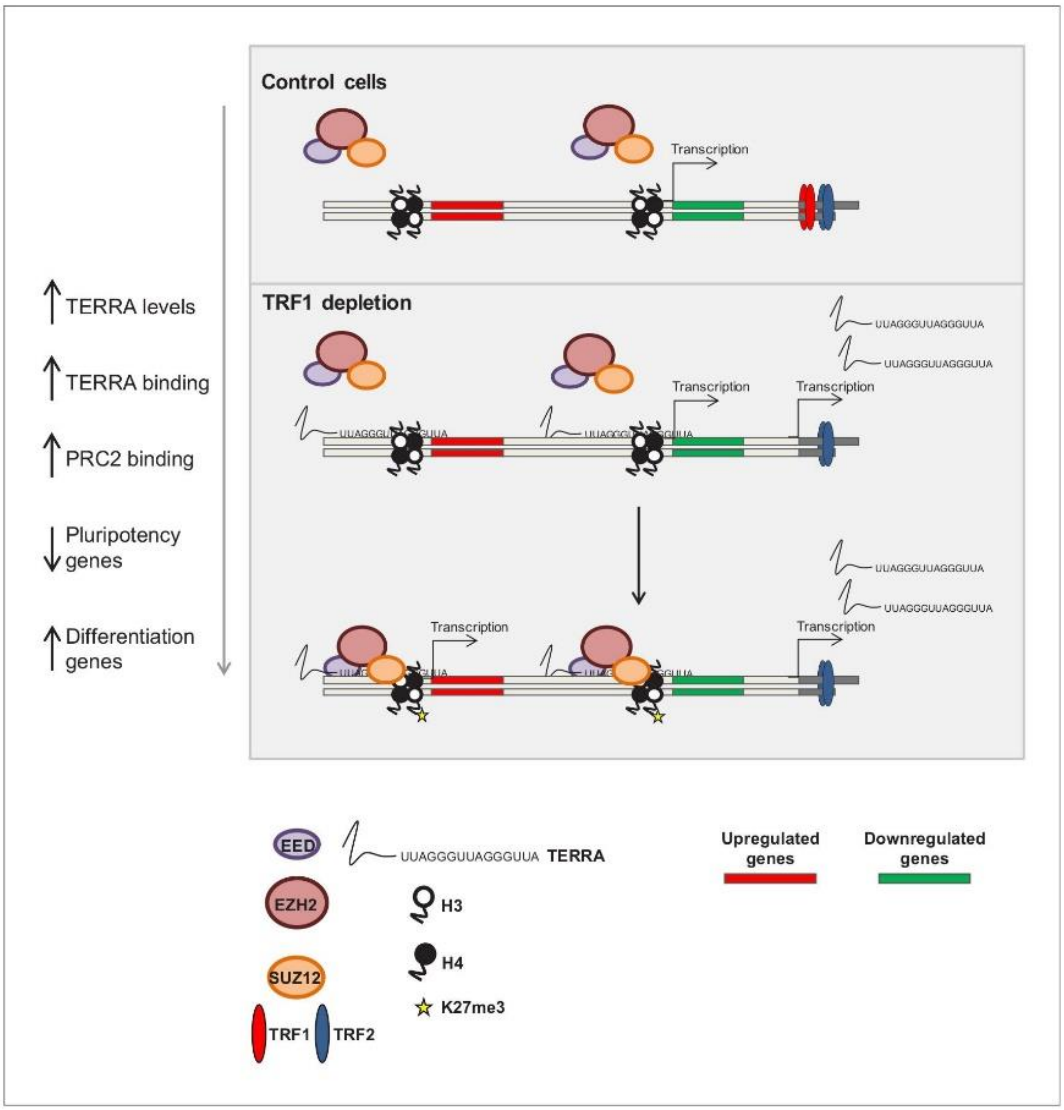


Figure 8. Model of the role of TRF1 in controlling pluripotency through TERRA expression and TERRA-dependent polycomb recruitment. In control 2i-grown iPS cells (top) the levels of TRF1 are elevated, the PRC2 complex is barely bound to the genome and pluripotency genes are expressed. When TRF1 levels are downregulated (bottom), TERRA expression is greatly increased, resulting in higher levels of binding to the genome. In this way, TERRA could increase PRC2 recruitment to many locations within the genome that are involved in the control of pluripotency and differentiation.
DOI: https://doi.org/10.7554/eLife.44656.020

(*Schneider et al., 2013*). In addition, we demonstrated that during in vivo reprogramming, TRF1 expression is highly upregulated in the de-differentiated areas upon reprogramming and that TRF1 chemical inhibition reduced the efficiency of in vivo reprogramming (*Marión et al., 2017*). TRF1 is also present at high levels in the Inner Cell Mass (ICM) of the blastocyst when compared to differentiated mouse embryonic fibroblasts (MEFs) (*Varela et al., 2011*). Together these findings suggested a role for TRF1 upregulation during in vitro and in vivo tissue reprogramming and pluripotency (*Marión et al., 2017*). However, the molecular mechanisms through which TRF1 modulates pluripotency has not yet been revealed.

Here, we make the unprecedented finding that TRF1 controls the maintenance of the pluripotency state by influencing the epigenetic status of the chromatin. In particular, TRF1 depletion in naïve-state pluripotent stem cells leads to dramatic changes in the expression of genes that are associated with pluripotency pathways, as well as of genes that are controlled by the polycomb complex PRC. These changes are consistent with loss of the naïve state as the PRC2 complex

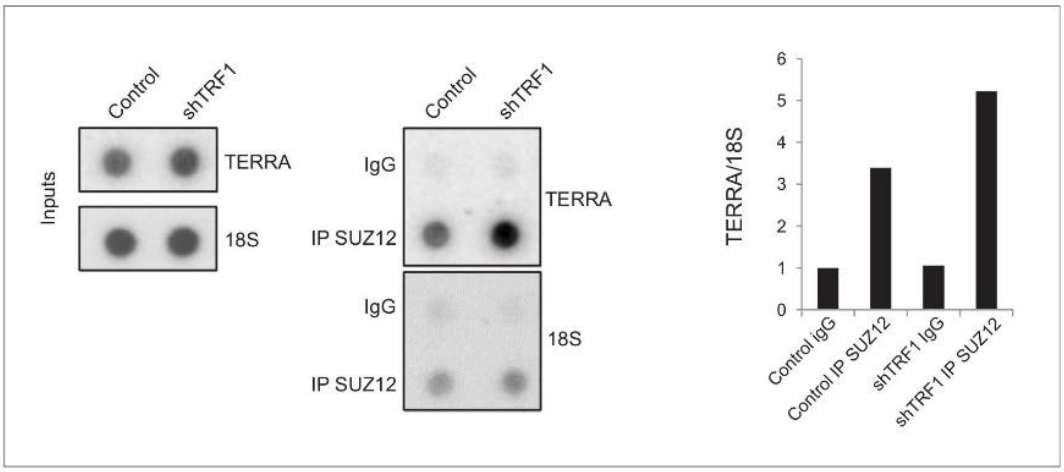


Figure 9. SUZ12 binding to TERRA increases upon TRF1 depletion. Immunoprecipitation of SUZ12 and detection of bound TERRA by dot-blot hybridization. Left, representative images of the input and immunoprecipitated TERRA in control cells and in cells depleted for TRF1. Note that levels of TERRA are higher in the SUZ12 immunocomplex when TRF1 is abrogated. Right, quantification of the TERRA and 18S levels in the immunocomplex. Levels of TERRA were normalized by the levels of 18S. Note again that binding of SUZ12 increases upon TRF1 depletion. The experiment was performed once.

DOI: https://doi.org/10.7554/eLife.44656.021

is recruited to bivalent genes and the differentiation programs are initiated. We further demonstrate that TRF1 depletion leads to increased binding of PRC2 complex to its target genes, which is concomitant with an increase in the abundance of the H3K27me3 polycomb mark at some of these sites. Accordingly, we see that TRF1 depletion induces changes in the expression of the genes in which these TRF1-dependent PRC2 sites are located.

In *Arabidopsis thaliana*, telomere-repeat binding factors (TRBs) similar to TRF1 have been reported to recruit PRC proteins to different promoters through a telobox motif (***Zhou et al., 2016b***; ***Zhou et al., 2018***). Although the binding of human TRF1 to extratelomeric sites had been reported (***Simonet et al., 2011***), we show here for the first time that mouse TRF1 can directly bind to a set of genes containing TTAGGG/AATCCC repeats in iPS cells in a naïve state. Interestingly, a significant percentage of these genes are targets of the pluripotency regulator ZFP322A, opening the possibility that TRF1 may also exert a direct effect on pluripotency by directly regulating the transcription of these genes. However, the vast epigenetic changes observed here as the consequence of TRF1 depletion, which involve the re-localization of PRC2 to many genomic sites, are not likely to be explained by the binding of TRF1 to these few extra-telomeric sites.

TRF1 has been previously shown to interact with the telomere-originated long non-coding RNAs known as TERRA (***Deng et al., 2009***; ***Chu et al., 2017***). We have previously shown that TERRA interacts with polycomb components and that this interaction is important to establish polycomb marks and heterochromatin marks at telomeric chromatin (***Montero et al., 2018***). Here, we demonstrate that TERRA can also bind to PRC2 sites in the genome that are dependent on TRF1. More importantly, we show that TRF1 depletion causes a dramatic increase in TERRA expression, in agreement with previous observations in human cells that show increased levels of TERRA upon TRF1 downregulation (***Porro et al., 2014***; ***Zeng et al., 2017***; ***Sadhukhan et al., 2018***). This upregulation of TERRA expression coincides with an increased binding of TERRA to genomic regions where PRC2 is also increased upon TRF1 depletion and that are important in controlling pluripotency and differentiation. Interestingly, these TERRA- and PRC2-bound regions also show altered expression upon TRF1 abrogation. Finally, we demonstrate that SUZ12 binding to TERRA increases upon TRF1 depletion, reinforcing the idea that TERRA mediates the recruitment of SUZ12 to the genome.

All together, these findings suggest a model by which TRF1-dependent TERRA upregulation allows TERRA-mediated recruitment of PRC2 to key pluripotency and differentiation genes, thus controlling the epigenetic landscape of pluripotent stem cells (see the model in ***Figure 8***). Finally, these findings explain the fact that TRF1 is upregulated during in vitro and in vivo reprogramming

and that it is essential for the reprogramming and maintenance of pluripotency. They also are in agreement with the fact that TERRA is tightly regulated during pluripotency and differentiation.

Materials and methods

Cells and culture conditions

The primary keratinocytes and iPS used in this work were generated in our lab and were tested for mycoplasm in a routine manner. *Trp53*$^{-/-}$, *Terf1*$^{+/+}$, *Terf1*$^{+/GFP}$ and *Terf1*$^{GFP/GFP}$ iPS cells were cultured in gelatin-coated plates with DMEM (high glucose) supplemented with serum replacement (KSR, Invitrogen), LIF 1000 u/ml, non-essential amino acids, glutamax and beta-mercaptoethanol (known as iPS medium). The medium was supplemented with MEK inhibitor PD0325901 (1 µM) and GSK3 inhibitor CHR99021 (3 µM), together known as 2i. Primary keratinocytes from newborn mice were obtained and grown as described in previous work (**Martínez et al., 2009**).

ES cells were plated in complete ES medium (FCS + LIF) on gelatin-coated 35mm plates at a density of 2.5×10^5 cells per plate. Next day, the medium was changed to ES medium (–LIF) containing 1 µM retinoic acid. Cells were cultured for 24, 48, 72 hr to 5 days changing medium daily.

TRF1 or TRF2 abrogation in iPS cells

Lentiviral supernatants were produced in HEK-293T cells (5×10^6 cells per 100-mm-diameter dish) transfected with the packaging plasmids pMDLg/pRRE (3.25 µg), pRSV.Rev (1.25 µg), pMDG VSVG (1.75 µg) (obtained from Addgene), and one of three shRNA lentiviral constructs (5 µg), pLKO.1-puro-scramble shRNA (obtained from Addgene), pLKO.1-puro-TRF1 shRNA (bacterial glicerol stock (TRCM0000071298, obtained from Sigma-Aldrich) or pLKO.1-puro-TRF2 shRNA (CCTTGGAA TCAGCTATCAATG). Transfections were performed using Fugene-6 transfection reagent (Roche) according to the manufacturer's protocol. Cells were cultured in standard ES cells medium containing LIF. After 2 days, viral supernatants (10 ml) were collected serially during the subsequent 36 hr, at 12 hr intervals, each time adding fresh ES cell medium to the cells (10 ml).

The recipient *Trp53*$^{-/-}$ iPS cells had been seeded the previous day (2×10^6 cells per 100-mm-diameter dish) in 2i-containing iPS medium, and received 4.5 ml of the corresponding viral supernatants (supplemented with 2i) and 5.5 ml of 2i-containing iPS medium. The procedure was repeated three times at 12 hr intervals. 12 hr after the last infection was completed, and the medium was replaced with 2i-containing iPS medium to remove the virus. The following day, the iPS cells were expanded in 2i-containing iPS medium supplemented with puromycin to select for the shRNAs vectors. After 48 hr, samples for the different experiments were collected.

Telomeric FISH

TRF1 was depleted as described above. Metaphase and Q-FISH hybridization was performed as previously described (**Samper et al., 2001**; **Gonzalo et al., 2006**).

Western blots

Whole-cell extracts were prepared by resuspending cells (10×10^6 cells/ml) in a buffer containing 10 mM Hepes (ph 7.9), 10 mM KCl, 2.5 mM $MgCl_2$, 0.34 M sucrose and 10% glycerol, supplemented with 0.1 mM phenylmethylsulfonyl fluoride (PMSF), protease inhibitors cocktail and Dnase (100 u/ml). After 30 min on ice, extracts were sonicated and resolved on NuPAGE 4–12% gradient Bis-Tris gels. After protein transfer onto nitrocellulose membrane (Whatman), the membranes were incubated with the primary indicated antibodies: rat antibody against TRF1 (raised in our laboratory against full-length mouse TRF1 protein) (1:500), mouse monoclonal against β-actin (A1978 Sigma, 1:5000), rabbit polyclonal antibody against SUZ12 (ab12073, 1:500), rabbit antibody against H3K27me3 (07–449 Upstate, 1:3000), mouse monoclonal antibody against phospho histone H2A.X (Ser139) (Millipore 05–636, 1:200) and rabbit antibody against SMC1 (A300-055A Bethyl, 1:1000). Antibody binding was detected after incubation with a secondary antibody coupled to horseradish peroxidase using chemiluminescence with ECL detection KIT (GE Healthcare). Western blots for SUZ12, H3K27me3 and H2A.X were performed with samples obtained from four independent TRF1 deletion experiments. Statistical analysis was performed by Student's t-test.

RNA-seq

For RNA-seq analysis, two independent TRF1 deletion experiments were carried out, obtaining two controls samples and two TRF1-deleted samples. Total RNA was extracted with TRIzol reagent (ThermoFisher 15596026) and then purified and treated with DNAse using an RNeasy Kit (QUIAGEN 74106), following the manufacturer's instructions. Total RNA (1 µg) was used for the RNA-seq experiment, and the obtained sample RNA Integrity Number was 10 (Agilent 2100 Bioanalyzer). The poly (A)+ fraction was purified and randomly fragmented, converted to double-stranded cDNA and processed through subsequent enzymatic treatments of end-repair, dA-tailing, and ligation to adapters with the 'NEBNext Ultra II Directional RNA Library Prep Kit for Illumina' (NEB, Cat. No. E7760) as recommended by the manufacturer. This kit incorporates dUTP during second-strand cDNA synthesis, which implies that only the cDNA strand generated during first strand synthesis is eventually sequenced. An adapter-ligated library was completed by PCR with Illumina PE primers. The resulting purified cDNA library was applied to an Illumina flow cell for cluster generation and sequenced on an Illumina NextSeq 500 sequencer, following manufacturer's protocols. Single-end sequenced reads were analyzed with the next*presso* pipeline (**Graña et al., 2018**) as follows. Sequencing quality was checked with FastQC v0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned to the mouse reference genome (NCBI37/mm9, https://ccb.jhu.edu/software/tophat/igenomes.shtml) with TopHat-2.0.10 (**Trapnell et al., 2012**) using Bowtie 1.0.0 (**Langmead et al., 2009**) and Samtools 0.1.1.9 (**Li et al., 2009**), allowing three mismatches and 20 multihits. Quantification of transcripts and differential expression were calculated with Cufflinks 2.2.1 (**Trapnell et al., 2012**), using the mouse NCBI37/mm9 transcript annotations from https://ccb.jhu.edu/software/tophat/igenomes.shtml. For differential expression, a false discover rate (FDR) <0.05 was used. GSEAPreranked (**Subramanian et al., 2005**) was used to perform gene set enrichment analysis on a pre-ranked gene list, setting 1000 gene set permutations. Only those gene sets with significant enrichment levels (FDR q-value <0.25) were finally considered.

ChIP-seq

The ChIP-seq experiment was performed twice for each antibody. *Trp53*$^{-/-}$ iPS cells and *Trp53*$^{-/-}$ iPS cells in which TRF1 expression had been abrogated (see above) were crosslinked with 1% formaldehyde (added to the medium) for 15 min at room temperature (RT). The crosslinking was stopped by adding 0.125M glycine to the medium for 5 min at RT. Fixed cells were washed with PBS containing 1 µM PMSF and protease inhibitors and then pelleted. Cells were lysed in lysis buffer (1% SDS, 10 mM EDTA and 50 mM Tris-HCl (pH 8.1)) at 2×10^7 cells/ml for 20 min at 4˚C. Sonication was performed with a Covaris system (shearing time 20 min, 20% duty cycle, intensity 6,200 cycles per burst, and 30 s per cycle). 50 µg of chromatin per immunoprecipitation reaction were diluted 1/10 in dilution buffer (1% Triton X-100, 2 mM EDTA (pH 8.0), 150 mM NaCl, 20 mM Tris-HCl (pH 8.1)) and immunoprecipitated with a rabbit sera against TRF1 (raised in our laboratory), rabbit polyclonal antibody against SUZ12 (ab12073) or rabbit antibody against H3K27me3 (07–449 Upstate), using protein A/G agarose beads (Santa Cruz Biotechnology). The beads were washed once with low salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl (pH 8.1), 150 mM NaCl), once with high-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl (pH 8.1), 500 mM NaCl), once with LiCl wash buffer (0.25 M LiCl, 1% NP40, 1% deoxycholateNa, 1 mM EDTA, 10 mM Tris-HCl (pH 8.1)), and twice with TE. The immune complexes were eluted with 500 µl of elution buffer (1% SDS, 0.1 M $NaHCO_3$). Reverse crosslinking was achieved through the addition of 20 µl of 5M NaCl and incubation at 65˚C for 8 hr. DNA was recovered by RNase and proteinase K treatment, phenol/chloroform extraction and ethanol precipitation. For each sample, 10–15 ng of DNA were used for the ChIP-seq experiments. Samples were processed through subsequent enzymatic treatments of end-repair, dA-tailing, and ligation to adapters with 'NEBNext Ultra II DNA Library Prep Kit for Illumina' from New England BioLabs (catalog # E7645). Adapter-ligated libraries were completed by limited-cycle PCR and extracted with a [single] double-sided size selection for library preparation (SPRI). Resulting average fragment size is 490 bp, from which 120 bp correspond to adaptor sequences. Libraries were applied to an Illumina flow cell for cluster generation and sequenced on an Illumina NextSeq 500 sequencer, by following the manufacturer's protocols.

CHIRT-seq and CHIRT-qPCR

CHIRT was performed once, mostly according to the protocol described by **Chu et al. (2017)** with some modifications. 30 millions cells were fixed and chromatin was fragmented in a Covaris system (shearing time 80 min, 20% duty cycle, intensity 6,200 cycles per burst, and 30 s per cycle). An antisense TERRA RNA biotinylated transcript consisting of eight CCCTAA repeats was used as bait for TERRA capture. After CHIRT, chromatin was eluted using RNAseH (NEB) as previously described (**Chu et al., 2017**), and the obtained DNA was used to prepare the libraries for deep sequencing. As input control, total fragmented chromatin was used. 0.5 ng of DNA per sample were used for the ChIRT-seq experiment (5 ng for the INPUT). Samples were processed through subsequent enzymatic treatments for fragmentation (5 min), end-repair, dA-tailing, and ligation to adapters with the 'NEBNext Ultra II FS DNA Library Prep Kit for Illumina' from New England BioLabs (catalog # E7805). Adapter-ligated libraries were completed by limited-cycle PCR and extracted with a [single] double-sided SPRI. Resulting average fragment size is 400 bp, from which 120 bp correspond to adaptor sequences. Libraries were applied to an Illumina flow cell for cluster generation and sequenced on an Illumina NextSeq 500 sequencer, following manufacturer's protocols. To determine changes in TERRA binding in the shTRF1 CHIRT sample compared to the control CHIRT sample, quantitative real-time PCR was performed, using the DNA libraries described above, by Go-Taq qPCR master mix (Promega) according to the manufacturer's protocol. All values were obtained in triplicates. In each case, values were normalized to input and to a genomic region where TERRA binding was not found in the CHIRT-seq. Primers are listed in **Table 1**.

ChIP-seq and CHIRT-seq analysis

ChIP-seq and ChIRT-seq sequenced reads were processed with the RUbioSeq pipeline (**Rubio-Camarillo et al., 2017**), which performed read alignment to the mouse reference genome (NCBI37/mm9, https://ccb.jhu.edu/software/tophat/igenomes.shtml) with BWA 0.7.10 (**Li and Durbin, 2009**), removed duplicated reads with the Picard tools v1.107 (https://broadinstitute.github.io/picard/) and converted the resulting alignment files to BED format with BEDTools 2.16.2 (**Quinlan and Hall, 2010**). All ChIP and input samples were randomly normalized to the same number of reads. Peak calling was performed with MACS2 v2.0.10.20130712 (**Quinlan and Hall, 2010**). Significant peaks were associated with nearby genes using Homer v4.10.1 (**Heinz et al., 2010**) and functional annotations for associated genes were obtained with Homer and Enrichr (**Kuleshov et al., 2016**). Heatmaps with signal density over called peaks and gene TSS sites were calculated and plotted with Deeptools (**Ramírez et al., 2016**).

Data set repository

Raw sequencing data and processed files for the RNA-seq, ChIP-seq and TERRA ChIRT-seq experiments are available from the Gene Expression Omnibus (GEO) under the ID GSE121759.

Quantitative PCR

Samples from three independent TRF1 deletion experiments were used for validation of RNA-seq results by quantitative PCR. Total RNA from cells was extracted as described above (see RNA-seq). Reverse transcription was performed using the iSCRIPT cDNA synthesis kit (BIO-RAD) according to manufacturer's protocols. Quantitative real-time PCR was performed using Go-Taq qPCR master mix (Promega) according to the manufacturer's protocol. All values were obtained in triplicates. Primers are listed in **Table 1**. Statistical analysis were performed using one tail, unpaired Student's t-test.

Preparation of nuclear extract, immunoprecipitation and mass spectrometry

The preparation of nuclear extract for immunoprecipitation and mass spectrometry was described in **Schneider et al. (2013)**.

TERRA RNA-FISH

Control *Trp53*$^{-/-}$ iPS cells and *Trp53*$^{-/-}$ iPS cells depleted for TRF1 (see above) from one single experiment were seeded in 96-well bottom-glass plates (Greiner Bio-One) coated with 0.1% porcine gelatine, and grown in iPS medium (see above) supplemented with MEK inhibitor PD0325901 (1 μM)

Table 1. List of oligos.

	Forward	Reverse
Actin	5′-GGCACCACACCTTCTACAATG-3′	5′-GTGGTGGTGAAGCTGTAG-3′
Terf1	5′-TCTAAGGATAGGCCAGATGCCA-3′	5′-CTGAAATCTGATGGAGCACGT-3′
Myc	5′-GTGCTGCATGAGGAGACACC-3′	5′-AGGGGTTTGCCTCTTCTCC-3′
Lama1	5′-CTGTCACCCTGGACTTACGG-3′	5′-GGTTTGAACTTGACGCCATC-3′
Id1	5′-CTGAACGGCGAGATCAGTG-3′	5′-GCCTCAGCGACACAAGATG-3′
Olig2	5′-CAGCGAGCACCTCAAATCTA-3′	5′-CCCCAGGGATGATCTAAGC-3′
Foxd3	5′-CTTTCTTTCGGGGGACACTC-3′	5′-CCAGGAGCGAGCAGAGAG-3′
Zfp423	5′-TGACGTTCGAGAACGAGAGAG-3′	5′-GGGAGTCGAACATCTGGTTG-3′
Pou2f3	5′-ATCGACGCCAAAAGGAGAAG-3′	5′-TGGACAGGAGGGACTGAGAG-3′
Sgk1	5′-AGAGGAGTCCTGTTCCTGGG-3′	5′-GGTCAGGATGTTGGCATGAT-3′
Igf2	5′-GCTTGTTGACACGCTTCAGTT-3′	5′-AAGCAGCACTCTTCCACGAT-3′
Bmp4	5′-CGCTTCTGCAGGAACCAA-3′	5′-ATCAAACTAGCATGGCTCGC-3′
Hoxc12	5′-ACCCTGGCTCTCTGGTTTC-3′	5′-CAACTTCGAATACGGCTTGC-3′
Jak3	5′-TTGGGGACTACTTGGCTGAG-3′	5′-AGAAGTCCTCAGTGGCCAGA-3′
Gbx2	5′-CTCGCTGCTCGCTTTCTCT-3′	5′-GGGTCATCTTCCACCTTTGA-3′
Myd88	5′-TATACTGAAGGAGCTGAAGTCGC-3′	5′-ACACTGCTTTCCACTCTGGC-3′
Khdc3	5′-GAATGCCTGGAAGATCCAAA-3′	5′-ATGTGGGATGTGCTCTCCAT-3′
Rbm47	5′-AAAGAACCAGGACCAATCGC-3′	5′-CACTGTTGGATCGCTGTTCA-3′
Zfp345	5′-TGGTCTTCCCAAACATAGCC-3′	5′-ACTTCACGTGGGAAGAGTGG-3′
Dnmt3b	5′-TCTAATGCCAAAGCTCACCC-3′	5′-CTCTTTGCCTCTCCAAGCTG-3′
Gata2	5′-CACCCCTAAGCAGAGAAGCA-3′	5′-CAGGCATTGCACAGGTAGTG-3′
Fzd5	5′-AGCAGGATCCTCCGAGAGTT-3′	5′-CAGCACTCAGTTCCACACCA-3′
Spry2	5′-GATTCAAGGGAGAGGGGTTG-3′	5′-CTCCATCAGGTCTTGGCAGT-3′
Zscan10	5′-GACGGAGAGGAGGTGGTACA-3′	5′-GCCAAGCTCTCTTCTCTGAGG-3′
Fgf17	5′-ATTGATTCTCTGCTGTCAAACACA-3′	5′-GCTGGTATTCACGGATTTGC-3′
Sall4	5′-TGCCTCGGTGTTAGATGTCA-3′	5′-GACAAAGGTGGGCTGTGCT-3′
Klf5	5′-GGATCTGGAGAAGCGACGTA-3′	5′-TCCTCAGGTGAGCTTTTAAGTGA-3′
Nr5a2	5′-TGGGAAGGAAGGGACAATCT-3′	5′-AACGCGACTTCTGTGTGTGA-3′
Tbx3	5′-CATCGCCGTTACTGCCTATC-3′	5′-GCCAGTGTCTCGAAAACCC-3′
Spp1	5′-TGACCCATCTCAGAAGCAGA-3′	5′-CATTGGAATTGCTTGGAAGAG-3′
Smad7	5′-CGAATTATCTGGCCCCTGG-3′	5′-GACACAGTAGAGCCTCCCCA-3′
Zbtb43	5′-GGTAGGCTGGAGCTACGGG-3′	5′-TGGCCATCAAAGAGCAGTC-3′
Adm	5′-CATCCAGCAGCTACCCTACG-3′	5′-TTCGCTCTGATTGCTGGCTT-3′
Cxcl12	5′-CACTCCAAACTGTGCCCTTC-3′	5′-AATTTCGGGTCAATGCACAC-3′
Bend6	5′-AGAAGCATCCGGAAGGAAAA-3′	5′-TGCCATTCCAACCAGTTCTT-3′
Fndc5	5′-GGTGCTGATCATTGTTGTGGT-3′	5′-CCTTGTTGTTATTGGGCTCG-3′
Fzd1	5′-GCTTACTCCTCAGCAGCACA-3′	5′-TCTCTCACCCATCCGTCAGT-3′
Fzd2	5′-CCTCAAGGTGCCGTCCTATC-3′	5′-GGATCCAGAGACGGGCAAAA-3′
Jun	5′-ACCGAGAATTCCGTGACGAC-3′	5′-TGAAAAGTCGCGGTCACTCA-3′
Proser2	5′-ACTTGAGCAGAGGTGGCAGT-3′	5′-GTGCTTCAGGCTCTCGTCAT-3′
Lgi2	5′-CCAAGGAGTCCATCATCTGC-3′	5′-CATTCGGTCCTTGATTTCCA-3′
Bmp7	5′-CCTGGGCTTACAGCTCTCTG-3′	5′-CCATGAAGGGTTGCTTGTTC-3′
Pitx2	5′-CAAAAAGGTCGAGTTCACGG-3′	5′-CTTTCCTTGCTGGCCCTTAT-3′
Foxa1	5′-AACAGCTACTACGCGGACAC-3′	5′-GCTCGTGGTCATGGTGTTCA-3′

Table 1 continued on next page

Table 1 continued

	Forward	Reverse
Otx2	5′-CTCCTGGAGGAGAGAGCAGTC-3′	5′-GGGTCCTTGGTGGGTAGATT-3′
Nkd2	5′-GGTGTGGAACATCGCTCAC-3′	5′-CTAGGGAACCCTTGTCGTCC-3′
Col2a1	5′-GCGGTCCTACGGTGTCAG-3′	5′-TTTATACCTCTGCCCATTCTGC-3′
Ppp2r2c	5′-GCCATCACTGATCGGAGC-3′	5′-AGACGAAGAGGTTGCAGTGG-3′
App	5′-CTCTACAATGTCCCTGCGGT-3′	5′-AGCTGATTCTGGGCTCACTG-3′
Satb1	5′-TATGAACCAGAGTTCGTTGGC-3′	5′-TTTGCTGCTGAGACATTTGC-3′
Nfix	5′-TCTGGCTTACTTTGTCCACACTC-3′	5′-GTTGGGCAGTGGTTTGATGT-3′
Foxn4	5′-ACCACTGCTCTCCACAGGAA-3′	5′-CAGGACAGCGACTGAAGGTC-3′
Trpv4	5′-ACCACCCCAGTGACAACAAG-3′	5′-ATGGGCCGATTGAAGACTTT-3′
Hspa1a	5′-TTTGTGTATTGCACGTGGGC-3′	5′-GGGGCAGTGCTGAATTGAAG-3′
Prkar2b	5′-AGGCTTGCAAAGACATCCTG-3′	5′-TGTTCCCCTTCTTTGACCAAT-3′
Lef1	5′-ACCCGTACATGTCAAATGGG-3′	5′-GTCGCTGTAGGTGATGAGGG-3′

CHIRT	Forward	Reverse
Chr18	5′-CAGCCTTTGTCCTTCACAGTT-3′	5′-GGTTCATAAGGCTTTTCTCCA-3′
ChrX	5′-TGTTCCCTCACAGCACAGAG-3′	5′-TAAGCCAGCCTCTCCAAAGA-3′
Pros1	5′-GGCAGTCTCTGGAGTTGGAA-3′	5′-CTAGCATCCCTTCCCCATTC-3′
Ddx58	5′-AAGTGGGGTTTCAGAGAGCA-3′	5′-CCCTAACCCTTCCCCATAAA-3′
Arid5b	5′-CTCTTCCCCTGGAGATCCTT-3′	5′-TTGGAAACAGATTTGAGCATTC-3′
Syde2	5′-GCTGGGTTTACCCCAATACA-3′	5′-GACCCACTTCCTAAGGACAGAA-3′
Zdbf2	5′-CATGGGGAAAGCATAATTGC-3′	5′-AGGCTTGGGACTCCTCTTGT-3′
Olig2	5′-CCACACCCTGTGTGTCTGTC-3′	5′-TCAACCTTCCGAACTTGAGG-3′
Lrrc8c	5′-GCTAGGTTCTGGGGACTGG-3′	5′-CAACCGCGTTTTCTCCTAGT-3′
Sorl1	5′-GCCTACCTCAGAATGGAGGTC-3′	5′-CACACACACACACTACCATATAATCC-3′
Mpdz	5′-GCGTCCCATCTTAAAACCAA-3′	5′-GATCCTCTCCATCCCTACCC-3′
Prmt8	5′-GGTTTGGGACTTAGGGGAAC-3′	5′-AGTTCCTTTCCCCCTTGAAA-3′
Dleu7	5′-TCAAGACTGGACCCCAAAAC-3′	5′-GGACCAGCCAGCTTGTATGT-3′
Dcc	5′-TTCAGTCCCTGGACAGACAG-3′	5′-ACACGCCTTTCCTTCACAGT-3′
Control region	5′-CAATGCCTAGATATACCGATCTCTT-3′	5′-CTCAGGACAAGACCCCACTG-3′

DOI: https://doi.org/10.7554/eLife.44656.022

and GSK3 inhibitor CHR99021 (3 μM). Next day, the cells were washed with PBS twice and incubated in cytobuffer (100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 10 mM pipes (pH 6.8)) for 30 s, incubated in cytobuffer with 0.5% Triton X-100 for 30 s, incubated in cytobuffer for 30 s, and fixed for 10 min in 4% paraformaldehyde in PBS. Fixed cells were dehydrated in 70% ethanol three times, once in 80%, 95%, and 100% ethanol, air-dried, and hybridized overnight at 37°C with a telomere-specific PNA-FITC probe (Panagene) in hybridization buffer (2 × sodium saline citrate (SSC)/50% formamide). Next day, cells were washed twice for 15 min in 2 x SSC, 50% formamide at 40°C, twice for 10 min in 2 × SSC at 40°C, for 10 min in 1 x SSC at 40°C, for 5 min in 4 x SSC at room temperature and for 5 min in 4 x SSC containing 0.1% Tween-20. The cells were incubated with DAPI (Molecular Probes) at room temperature for 10 min and washed three times with PBS. Fluorescence signal was preserved in Vectashild (Vector laboratories). Signals were visualized in a confocal ultra spectral microscope SP5-WLL (Leica). Statistical analyses were performed using Student's t-test, where n indicates the number of cells of each sample analyzed.

TERRA northern blot

Northern blot analyses were performed using standard protocols. Telomere probe was obtained by excising a 1.6 Kb $(TTAGGG)_n$ DNA insert from pNYH3 plasmid (a kind gift from T. de Lange, Rockefeller University, NY, USA). Northern blots were normalized using 18S or 28 s probes and quantified using ImageJ. The probes were labeled using the commercial Prime-It II Random Primer Labeling Kit (Agilent Genomics). For TRF1 deletion in iPS cells, samples were obtained from three independent TRF1 deletion experiments. Statistical analysis was performed using Student's t-test. For TERRA analysis in primary keratinocytes from newborn mice, three newborn mice from each genotype were used. Statistical analysis was performed using Student's t-test.

Immunofluorescence

2i-grown *Trp53*$^{-/-}$ iPS cells and 2i-grown *Trp53*$^{-/-}$ iPS cells where TRF1 expression had been abrogated, or ES cells treated or untreated with retinoic acid, were plated in gelatin-coated coverslips, incubated with cytobuffer (100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 10 mM pipes (pH 6.8)) with 0.5% Triton X-100 for 6 min, and fixed for 10 min in 4% paraformaldehyde in PBS. Cells were then blocked in 10% BSA in PBS for 1 hr at RT and incubated with the primary antibodies dissolved in Dako antibody diluents (Dako) for 1 hr in a humid chamber at RT. The following antibodies were used: rat antibody anti-TRF1 (home made, 1:250), rabbit antibody against RAP1 (A300-306A, Bethyl, 1:250) and mouse monoclonal antibody against phospho histone H2A.X (Ser139) (Millipore 05–636, 1:250). Cells were washed 3 times for 30 min with PBS containing 0.1% Tween-20 and incubated with Alexa secondary antibodies for 1 hr at RT in a humid chamber. After 3 washes of 30 min with PBS containing 0.1% Tween-20, samples were mounted in Prolong with DAPI. Fluorescent signals were visualized in a confocal ultraspectral microscope SP5-WLL (Leica). The experiment was performed once. Statistical analysis were performed using Student's t-test, where n indicates the number of cells in each analyzed sample.

Enrichr analysis

The p-values shown in all the figures containing Enrichr analysis (Wikipathways, KEGG Pathways, ENCODE histone modifications, CHEA) were the adjusted p-values provided by Enrichr.

RNA immunoprecipitation (RIP)

Trp53$^{-/-}$ 2i-grown iPS cells, either controls or cells in which TRF1 was depleted) were crosslinked with 1% formaldehyde for 15 min and the quenched with glycine 125 mM for 5 min. Cells were washed with PBS, scraped from the plate and pelleted for 5 min at 1200 rpm. 25×10^6 cells per immunoprecipitation were resuspended in 7 ml of ice-cold cell lysis buffer (5 mM Tris (pH 8), 85 mM KCl, 0.5% Igepal), and incubated for 10 min at 4˚C in rotation. Crude nuclear sample was collected by centrifugation for 10 min at 3000 g at 4˚C. The nuclear pellet was washed with 0.5 ml of cell lysis buffer without Igepal (5 mM Tris (pH 8), 85 mM KCl) and microcentrifuged for 10 min at 3000 g at 4˚C. The nuclear pellet was resuspended in 1 ml of RIPA buffer (150 mM NaCl, 50 mM Tris (pH 7.4), 0.1% SDS, 0.5% Na-Doc, 1% Triton +0.1M DTT, RNAsin, Proteinase inhibitor 1x) to lyse the nucleus and incubated in ice for 10 min. The sample was doused 20 times and sonicated thrice for 5 s on setting 3. The sonicated lysate was microcentrifuged for 10 min at 4˚C at 14,000 rpm. A fraction of the soluble chromatin (supernatant) was kept as input. The sample was pre-cleared with Protein A+G beads for 1 hr at 4˚C and centrifuged at 2000 rpm for 4 min. The precleared sample was then incubated o/n with the corresponding antibody (IgG as negative control or antibody anti SUZ12 (ab12073)). After o/n incubation, the beads were washed once with low-salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris (pH 8), 150 mM NaCl, +0.1M DTT, RNAsin, Proteinase inhibitor 1x), once with high-salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris (pH 8), 500 mM NaCl, +0.1M DTT, RNAsin, Proteinase inhibitor 1x), once with LiCl buffer (0.25 M LiCl, 1% Igepal, 1% deoxycholate, 1 mM EDTA, 10 mM Tris (pH 8), +0.1M DTT, RNAsin, Proteinase inhibitor 1x), and twice with TE (pH 8) (100 mM Tris (pH 8), 10 mM EDTA (pH 8)). Each wash was performed at 4˚C for 5 min in rotation. The RNA was eluted with 250 ul of elution buffer (25 mM Tris (pH 7.5), 5 mM EDTA, 0.5% SDS), treated with proteinase K for 45 min at 45˚C and incubated 2 hr at 65˚C to reverse crosslink. RNA extracted with trizol and purified and DNAse-treated on using RNeasy Kit (QUIAGEN 74106),

following the manufacturer's instructions. The RNA was loaded using a dot-blot and the presence of TERRA and 18S RNAs was detected as described before (***Marion et al., 2009***).

Acknowledgements

We thank Ana Cuadrado for her help with the setting up of ChIP-seq. We thank Tommaso Vicanolo for his help in differentiating cells. We are indebted to Orlando Domínguez and the Genomic Unit from CNIO for their advice and for the generation of the libraries. We thank Diego Megías from the Confocal Microscopy Unit at CNIO for his help with the quantification of the immunofluorescence and RNA-FISH. We are grateful to Jeannie T Lee and Catherine Hsueh-Ping Chu for kindly sharing the CHIRT protocol and for assistance with details. Research in the Blasco lab is funded by the Spanish Ministry of Economy and Competitiveness Projects (SAF2013-45111-R and SAF2015-72455-EXP), the Comunidad de Madrid Project (S2017/BMD-3770), the World Cancer Research (WCR) Project (16–1177) and the Fundación Botín (Spain).

Additional information

Competing interests

José Alejandro Palacios-Fábrega: The author is affiliated with Astellas Pharma Europe Ltd. The author has no other competing interests to declare. The other authors declare that no competing interests exist.

Funding

Funder	Grant reference number	Author
Ministerio de Economía y Competitividad	SAF2013-45111-R	Isabel López de Silanes
Consejería de Educación, Juventud y Deporte, Comunidad de Madrid	S2017/BMD-3770	Juan J Montero Maria A Blasco
World Cancer Research Fund	16-1177	Maria A Blasco
Fundación Botín		Maria A Blasco
Ministerio de Economía y Competitividad	SAF2015-72455-EXP	Maria A Blasco

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Rosa María Marión, Formal analysis, Designed, performed and interpreted most of the experiments, Writing – original draft, Writing – review and editing; Juan J Montero, Designed, performed and interpreted most of the experiments; Isabel López de Silanes, Performed the differentiation experiment; Osvaldo Graña-Castro, Performed the bioinformatic analysis; Paula Martínez, Stefan Schoeftner, Provided the keratinocytes, Performed Northern Blots; José Alejandro Palacios-Fábrega, Performed the proteomics analysis; Maria A Blasco, Secured funding, Conceived the original idea, Designed and interpreted results, Writing – original draft, Writing – review and editing

Author ORCIDs

Rosa María Marión https://orcid.org/0000-0002-0374-3884
Maria A Blasco https://orcid.org/0000-0002-4211-233X

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.44656.030
Author response https://doi.org/10.7554/eLife.44656.031

Additional files

Supplementary files

- Supplementary file 1. Table S1.

DOI: https://doi.org/10.7554/eLife.44656.023

- Supplementary file 2. Table S2.

DOI: https://doi.org/10.7554/eLife.44656.024

- Supplementary file 3. Table S3.

DOI: https://doi.org/10.7554/eLife.44656.025

- Supplementary file 4. TRF1 does not interact with SUZ12. Mass spectroscopy of protein complexes immunoprecipitated with an anti-GFP antibody in *Terf1*$^{GFP/GFP}$, *Terf1*$^{+/GFP}$ and *Terf1*$^{+/+}$ iPS cells. The table shows the 64 proteins that showed co-immunoprecipitation in both *Terf1* $^{GFP/GFP}$ and *Terf1*$^{+/GFP}$ cells but not in *Terf1*$^{+/+}$ cells. Note the presence in this group of proteins of different components of shelterin, namely TPP1 (ACD), RAP1 (TERF2IP), TRF1 (TERF1), TIN2 (TINF2) and POT1, but not the presence of SUZ12. Related to ***Figure 2***.

DOI: https://doi.org/10.7554/eLife.44656.010

- Transparent reporting form

DOI: https://doi.org/10.7554/eLife.44656.026

Data availability

Raw sequencing data and additional processed files for the RNA-seq, ChIP-seq and TERRA CHIRT-seq experiments have been placed in the Gene Expression Omnibus (GEO) under the accession number GSE121759.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Marión RM, Montero JJ, Graña-Castro O, Blasco MA	2018	RNA-seq, ChIP-seq and TERRA CHIRT-seq from p53-/- iPS infected with a lentiviral virus carrying a control scrambled shRNA or shRNA against TRF1	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE121759	NCBI Gene Expression Omnibus, GSE121759

References

Azuara V, Perry P, Sauer S, Spivakov M, Jørgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, Fisher AG. 2006. Chromatin signatures of pluripotent cell lines. *Nature Cell Biology* **8**:532–538. DOI: https://doi.org/10.1038/ncb1403, PMID: 16570078

Benetti R, García-Cao M, Blasco MA. 2007. Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. *Nature Genetics* **39**:243–250. DOI: https://doi.org/10.1038/ng1952, PMID: 17237781

Blackburn EH. 2005. Telomeres and telomerase: their mechanisms of action and the effects of altering their functions. *FEBS Letters* **579**:859–862. DOI: https://doi.org/10.1016/j.febslet.2004.11.036, PMID: 15680963

Blasco MA. 2007. The epigenetic regulation of mammalian telomeres. *Nature Reviews Genetics* **8**:299–309. DOI: https://doi.org/10.1038/nrg2047, PMID: 17363977

Boué S, Paramonov I, Barrero MJ, Izpisúa Belmonte JC. 2010. Analysis of human and mouse reprogramming of somatic cells to induced pluripotent stem cells. What is in the plate? *PLOS ONE* **5**:e12664. DOI: https://doi.org/10.1371/journal.pone.0012664, PMID: 20862250

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**:947–956. DOI: https://doi.org/10.1016/j.cell.2005.08.020, PMID: 16153702

Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, Bell GW, Otte AP, Vidal M, Gifford DK, Young RA, Jaenisch R. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**:349–353. DOI: https://doi.org/10.1038/nature04733, PMID: 16625203

Chu HP, Cifuentes-Rojas C, Kesner B, Aeby E, Lee HG, Wei C, Oh HJ, Boukhali M, Haas W, Lee JT. 2017. TERRA RNA antagonizes ATRX and protects telomeres. *Cell* **170**:86–101. DOI: https://doi.org/10.1016/j.cell.2017.06.017, PMID: 28666128

de Lange T. 2005. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & Development* **19**:2100–2110. DOI: https://doi.org/10.1101/gad.1346005, PMID: 16166375

Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM. 2009. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Molecular Cell* **35**:403–413. DOI: https://doi.org/10.1016/j.molcel.2009.06.025, PMID: 19716786

Endoh M, Endo TA, Endoh T, Fujimura Y, Ohara O, Toyoda T, Otte AP, Okano M, Brockdorff N, Vidal M, Koseki H. 2008. Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135**:1513–1524. DOI: https://doi.org/10.1242/dev.014340, PMID: 18339675

Faust C, Schumacher A, Holdener B, Magnuson T. 1995. The eed mutation disrupts anterior mesoderm production in mice. *Development* **121**:273–285. PMID: 7768172

García-Cao M, Gonzalo S, Dean D, Blasco MA. 2002. A role for the rb family of proteins in controlling telomere length. *Nature Genetics* **32**:415–419. DOI: https://doi.org/10.1038/ng1011, PMID: 12379853

García-Cao M, O'Sullivan R, Peters AH, Jenuwein T, Blasco MA. 2004. Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nature Genetics* **36**:94–99. DOI: https://doi.org/10.1038/ng1278, PMID: 14702045

Gonzalo S, García-Cao M, Fraga MF, Schotta G, Peters AH, Cotter SE, Eguía R, Dean DC, Esteller M, Jenuwein T, Blasco MA. 2005. Role of the RB1 family in stabilizing histone methylation at constitutive heterochromatin. *Nature Cell Biology* **7**:420–428. DOI: https://doi.org/10.1038/ncb1235, PMID: 15750587

Gonzalo S, Jaco I, Fraga MF, Chen T, Li E, Esteller M, Blasco MA. 2006. DNA methyltransferases control telomere length and telomere recombination in mammalian cells. *Nature Cell Biology* **8**:416–424. DOI: https://doi.org/10.1038/ncb1386, PMID: 16565708

Graña O, Rubio-Camarillo M, Fdez-Riverola F, Pisano DG, Glez-Peña D. 2018. Nextpresso: next generation sequencing expression analysis pipeline. *Current Bioinformatics* **13**:583–591. DOI: https://doi.org/10.2174/1574893612666170810153850

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**:576–589. DOI: https://doi.org/10.1016/j.molcel.2010.05.004, PMID: 20513432

Karlseder J, Kachatrian L, Takai H, Mercer K, Hingorani S, Jacks T, de Lange T. 2003. Targeted deletion reveals an essential function for the telomere length regulator Trf1. *Molecular and Cellular Biology* **23**:6533–6541. DOI: https://doi.org/10.1128/MCB.23.18.6533-6541.2003, PMID: 12944479

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research* **44**:W90–W97. DOI: https://doi.org/10.1093/nar/gkw377, PMID: 27141961

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**:R25. DOI: https://doi.org/10.1186/gb-2009-10-3-r25, PMID: 19261174

Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K, et al. 2006. Control of developmental regulators by polycomb in human embryonic stem cells. *Cell* **125**:301–313. DOI: https://doi.org/10.1016/j.cell.2006.02.043, PMID: 16630818

Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K, Wutz A. 2010. Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes & Development* **24**:265–276. DOI: https://doi.org/10.1101/gad.544410, PMID: 20123906

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. DOI: https://doi.org/10.1093/bioinformatics/btp352, PMID: 19505943

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. DOI: https://doi.org/10.1093/bioinformatics/btp324, PMID: 19451168

López de Silanes I, Graña O, De Bonis ML, Dominguez O, Pisano DG, Blasco MA. 2014. Identification of TERRA locus unveils a telomere protection role through association to nearly all chromosomes. *Nature Communications* **5**:4723. DOI: https://doi.org/10.1038/ncomms5723, PMID: 25182072

Lund AH, van Lohuizen M. 2004. Polycomb complexes and silencing mechanisms. *Current Opinion in Cell Biology* **16**:239–246. DOI: https://doi.org/10.1016/j.ceb.2004.03.010, PMID: 15145347

Ma H, Ng HM, Teh X, Li H, Lee YH, Chong YM, Loh YH, Collins JJ, Feng B, Yang H, Wu Q. 2014. Zfp322a regulates mouse ES cell pluripotency and enhances reprogramming efficiency. *PLOS Genetics* **10**:e1004038. DOI: https://doi.org/10.1371/journal.pgen.1004038, PMID: 24550733

Marion RM, Strati K, Li H, Tejera A, Schoeftner S, Ortega S, Serrano M, Blasco MA. 2009. Telomeres acquire embryonic stem cell characteristics in induced pluripotent stem cells. *Cell Stem Cell* **4**:141–154. DOI: https://doi.org/10.1016/j.stem.2008.12.010, PMID: 19200803

Marión RM, López de Silanes I, Mosteiro L, Gamache B, Abad M, Guerra C, Megías D, Serrano M, Blasco MA. 2017. Common telomere changes during in Vivo Reprogramming and Early Stages of Tumorigenesis. *Stem Cell Reports* **8**:460–475. DOI: https://doi.org/10.1016/j.stemcr.2017.01.001, PMID: 28162998

Marks H, Kalkan T, Menafra R, Denissov S, Jones K, Hofemeister H, Nichols J, Kranz A, Stewart AF, Smith A, Stunnenberg HG. 2012. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**:590–604. DOI: https://doi.org/10.1016/j.cell.2012.03.026, PMID: 22541430

Martínez P, Thanasoula M, Muñoz P, Liao C, Tejera A, McNees C, Flores JM, Fernández-Capetillo O, Tarsounas M, Blasco MA. 2009. Increased telomere fragility and fusions resulting from TRF1 deficiency lead to degenerative pathologies and increased Cancer in mice. *Genes & Development* **23**:2060–2075. DOI: https://doi.org/10.1101/gad.543509, PMID: 19679647

Martínez P, Blasco MA. 2011. Telomeric and extra-telomeric roles for telomerase and the telomere-binding proteins. *Nature Reviews Cancer* **11**:161–176. DOI: https://doi.org/10.1038/nrc3025, PMID: 21346783

Montero JJ, López-Silanes I, Megías D, F Fraga M, Castells-García Á, Blasco MA. 2018. TERRA recruitment of polycomb to telomeres is essential for histone trymethylation marks at Telomeric heterochromatin. *Nature Communications* **9**:1548. DOI: https://doi.org/10.1038/s41467-018-03916-3, PMID: 29670078

O'Carroll D, Erhardt S, Pagani M, Barton SC, Surani MA, Jenuwein T. 2001. The polycomb-group gene Ezh2 is required for early mouse development. *Molecular and Cellular Biology* **21**:4330–4336. DOI: https://doi.org/10.1128/MCB.21.13.4330-4336.2001, PMID: 11390661

Pasini D, Bracken AP, Jensen MR, Lazzerini Denchi E, Helin K. 2004. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *The EMBO Journal* **23**:4061–4071. DOI: https://doi.org/10.1038/sj.emboj.7600402, PMID: 15385962

Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K. 2007. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Molecular and Cellular Biology* **27**:3769–3779. DOI: https://doi.org/10.1128/MCB.01432-06, PMID: 17339329

Pereira CF, Piccolo FM, Tsubouchi T, Sauer S, Ryan NK, Bruno L, Landeira D, Santos J, Banito A, Gil J, Koseki H, Merkenschlager M, Fisher AG. 2010. ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. *Cell Stem Cell* **6**:547–556. DOI: https://doi.org/10.1016/j.stem.2010.04.013, PMID: 20569692

Porro A, Feuerhahn S, Delafontaine J, Riethman H, Rougemont J, Lingner J. 2014. Functional characterization of the TERRA transcriptome at damaged telomeres. *Nature Communications* **5**:5379. DOI: https://doi.org/10.1038/ncomms6379, PMID: 25359189

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. DOI: https://doi.org/10.1093/bioinformatics/btq033, PMID: 20110278

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**:W160–W165. DOI: https://doi.org/10.1093/nar/gkw257, PMID: 27079975

Ringrose L, Paro R. 2004. Epigenetic regulation of cellular memory by the polycomb and trithorax group proteins. *Annual Review of Genetics* **38**:413–443. DOI: https://doi.org/10.1146/annurev.genet.38.072902.091907, PMID: 15568982

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**:1311–1323. DOI: https://doi.org/10.1016/j.cell.2007.05.022, PMID: 17604720

Rossiello F, Aguado J, Sepe S, Iannelli F, Nguyen Q, Pitchiaya S, Carninci P, d'Adda di Fagagna F. 2017. DNA damage response inhibition at dysfunctional telomeres by modulation of telomeric DNA damage response RNAs. *Nature Communications* **8**:13980. DOI: https://doi.org/10.1038/ncomms13980, PMID: 28239143

Rubio-Camarillo M, López-Fernández H, Gómez-López G, Carro Á, Fernández JM, Torre CF, Fdez-Riverola F, Glez-Peña D. 2017. RUbioSeq+: a multiplatform application that executes parallelized pipelines to analyse next-generation sequencing data. *Computer Methods and Programs in Biomedicine* **138**:73–81. DOI: https://doi.org/10.1016/j.cmpb.2016.10.008, PMID: 27886717

Sadhukhan R, Chowdhury P, Ghosh S, Ghosh U. 2018. Expression of Telomere-Associated proteins is interdependent to stabilize native telomere structure and telomere dysfunction by G-Quadruplex ligand causes TERRA upregulation. *Cell Biochemistry and Biophysics* **76**:311–319. DOI: https://doi.org/10.1007/s12013-017-0835-0, PMID: 29134494

Samper E, Flores JM, Blasco MA. 2001. Restoration of telomerase activity rescues chromosomal instability and premature aging in Terc-/- mice with short telomeres. *EMBO Reports* **2**:800–807. DOI: https://doi.org/10.1093/embo-reports/kve174, PMID: 11520856

Schneider RP, Garrobo I, Foronda M, Palacios JA, Marión RM, Flores I, Ortega S, Blasco MA. 2013. TRF1 is a stem cell marker and is essential for the generation of induced pluripotent stem cells. *Nature Communications* **4**:1946. DOI: https://doi.org/10.1038/ncomms2946, PMID: 23735977

Schoeftner S, Blasco MA. 2008. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nature Cell Biology* **10**:228–236. DOI: https://doi.org/10.1038/ncb1685, PMID: 18157120

Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL, de Lange T. 2009. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* **138**:90–103. DOI: https://doi.org/10.1016/j.cell.2009.06.021, PMID: 19596237

Shan Y, Liang Z, Xing Q, Zhang T, Wang B, Tian S, Huang W, Zhang Y, Yao J, Zhu Y, Huang K, Liu Y, Wang X, Chen Q, Zhang J, Shang B, Li S, Shi X, Liao B, Zhang C, et al. 2017. PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naïve ESCs. *Nature Communications* **8**:672. DOI: https://doi.org/10.1038/s41467-017-00668-4, PMID: 28939884

Simon JA, Kingston RE. 2009. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nature Reviews Molecular Cell Biology* **10**:697–708. DOI: https://doi.org/10.1038/nrm2763, PMID: 19738629

Simonet T, Zaragosi LE, Philippe C, Lebrigand K, Schouteden C, Augereau A, Bauwens S, Ye J, Santagostino M, Giulotto E, Magdinier F, Horard B, Barbry P, Waldmann R, Gilson E. 2011. The human TTAGGG repeat factors

1 and 2 bind to a subset of interstitial telomeric sequences and satellite repeats. *Cell Research* **21**:1028–1038. DOI: https://doi.org/10.1038/cr.2011.40, PMID: 21423270

Sparmann A, van Lohuizen M. 2006. Polycomb silencers control cell fate, development and Cancer. *Nature Reviews Cancer* **6**:846–856. DOI: https://doi.org/10.1038/nrc1991, PMID: 17060944

Stock JK, Giadrossi S, Casanova M, Brookes E, Vidal M, Koseki H, Brockdorff N, Fisher AG, Pombo A. 2007. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nature Cell Biology* **9**:1428–1435. DOI: https://doi.org/10.1038/ncb1663, PMID: 18037880

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* **102**:15545–15550. DOI: https://doi.org/10.1073/pnas.0506580102, PMID: 16199517

Theunissen TW, Jaenisch R. 2014. Molecular control of induced pluripotency. *Cell Stem Cell* **14**:720–734. DOI: https://doi.org/10.1016/j.stem.2014.05.002, PMID: 24905163

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nature Protocols* **7**:562–578. DOI: https://doi.org/10.1038/nprot.2012.016, PMID: 22383036

Varela E, Schneider RP, Ortega S, Blasco MA. 2011. Different telomere-length dynamics at the inner cell mass versus established embryonic stem (ES) cells. *PNAS* **108**:15207–15212. DOI: https://doi.org/10.1073/pnas.1105414108, PMID: 21873233

Voigt P, Tee WW, Reinberg D. 2013. A double take on bivalent promoters. *Genes & Development* **27**:1318–1338. DOI: https://doi.org/10.1101/gad.219626.113, PMID: 23788621

Wang X, Goodrich KJ, Gooding AR, Naeem H, Archer S, Paucek RD, Youmans DT, Cech TR, Davidovich C. 2017. Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. *Molecular Cell* **65**:1056–1067. DOI: https://doi.org/10.1016/j.molcel.2017.02.003

Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou M-M. 2010. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Molecular Cell* **38**:662–674. DOI: https://doi.org/10.1016/j.molcel.2010.03.021

Zeng S, Liu L, Sun Y, Lu G, Lin G. 2017. Role of Telomeric repeat-containing RNA in Telomeric chromatin remodeling during the early expansion of human embryonic stem cells. *The FASEB Journal* **31**:4783–4795. DOI: https://doi.org/10.1096/fj.201600939RR, PMID: 28765174

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**:750–756. DOI: https://doi.org/10.1126/science.1163045, PMID: 18974356

Zhou N, Li Q, Lin X, Hu N, Liao JY, Lin LB, Zhao C, Hu ZM, Liang X, Xu W, Chen H, Huang W. 2016a. BMP2 induces chondrogenic differentiation, osteogenic differentiation and endochondral ossification in stem cells. *Cell and Tissue Research* **366**:101–111. DOI: https://doi.org/10.1007/s00441-016-2403-0, PMID: 27083447

Zhou Y, Hartwig B, James GV, Schneeberger K, Turck F. 2016b. Complementary activities of TELOMERE REPEAT BINDING proteins and polycomb group complexes in transcriptional regulation of target genes. *The Plant Cell* **28**:87–101. DOI: https://doi.org/10.1105/tpc.15.00787, PMID: 26721861

Zhou Y, Wang Y, Krause K, Yang T, Dongus JA, Zhang Y, Turck F. 2018. Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in Arabidopsis. *Nature Genetics* **50**:638–644. DOI: https://doi.org/10.1038/s41588-018-0109-9, PMID: 29700471

Discusión

Discusión

1. Estudio del papel de los Telomeric repeat-containing RNA (TERRA) en la biología del telómero

Los telómeros son estructuras altamente compactadas al final de los cromosomas que evitan que este sea reconocido como daño de doble cadena, favoreciendo la estabilidad génica [77–81]. Al estar altamente heterocromatinizados y a que esa heterocromatinización se extiende hasta los subtelómeros, causando el llamado efecto de la posición telomérico o *telomere position effect* (TPE) [77,103], fenómeno por el cual se silencian los genes cercanos al mismo. Por ello, tradicionalmente se pensaba que eran transcripcionalmente inactivos. Sin embargo nuestro laboratorio y otros co-descubrieron hace ya una década la existencia de unos lncRNAs denominados *Telomeric repeat-containg RNA* o TERRA enriquecidos en la repetición telomérica UUAGGG que se transcriben desde el subtelómero hasta el telómero de las células de mamífero [151,152].

A pesar del tiempo transcurrido desde el descubrimiento de la transcripción telomérica, las funciones de TERRA no están del todo claras y presentan contradicciones entre sí. La falta de certeza en las funciones de TERRA se debe principalmente a que hasta ahora se desconocía los orígenes transcripcionales de TERRA, por lo que no se han podido generar modelos KO de TERRA. Por ello, el primer gran objetivo de esta tesis ha sido localizar el *locus* humano de TERRA y generar modelos TERRA KO. Para ello en el Artículo 1 caracterizamos si los 18 *loci* previamente propuestos como inicios transcripcionales de TERRA presentaban características propias de TERRA [160]. Solo encontramos características propias de TERRA en los transcritos procedentes de los subtelómero 20q y Xp. La eliminación del *loci* Xp no afectaba a los niveles de TERRA, mientras que las células 20q-TERRA KO presentaban una drástica disminución de los TERRA totales. Los principales fenotipos que observamos en las células 20q-TERRA KO fueron una pérdida de la longitud telomérica y un incremento del daño en DNA total y en DNA telomérico.

1.1. Limitaciones metodológicas para el estudio de TERRA debido a la falta de modelos *knock-out*

El principal motivo que ha dificultado la generación de modelos genéticos KO para TERRA es el desconocimiento de su origen transcripcional debido a la naturaleza repetitiva de TERRA. Hasta ahora, se desconocía si uno o varios telómeros eran los responsables del origen de TERRA. Esto a su vez ha llevado a que, en la mayoría de los estudios, se hayan centrado en la sobreexpresión, eliminación y detección de la repetición UUAGGG de TERRA sin prestar atención a su secuencia o su localización subtelomérica, más específica. Una de las principales consecuencias de esto es la dificultad y la variabilidad en las técnicas usadas para medir los niveles totales de TERRA. Desde su descubrimiento, los niveles TERRA se han medido prestando atención a la repetición telomérica, mediante Northern usando sondas teloméricas [151,152], o posteriormente mediante RNA-FISH usando de nuevo sondas contra la repetición telomérica [167,233]. Estas sondas presentan cierta inespecificidad por dos motivos. Estas sondas

tienen reactividad cruzada con los RNA ribosomales y además las secuencia telomérica TTAGGG no se encuentra limitada al telómero, sino que se encuentra distribuida por todo el genoma formando las llamadas secuencias teloméricas intersticiales o *Interstitial Telomeric Sequences* (ITSs) [32,33,34]. Para medir los niveles de TERRA también se han usado primers de qPCR específicos para subtelómeros concretos [156,160], pero al desconocerse el origen transcripcional de TERRA y al no haberse caracterizado estos subtelómeros como *bona fide* TERRA, cabe la posibilidad de que la transcripción en estos subtelómeros no llega al telómero y por lo tanto que estos RNAs no sean TERRA.

La falta de modelos TERRA KO ha originado una gran heterogeneidad de los modelos usados para el estudio de TERRA, tanto de los organismos usado para el mismo (levaduras, células humanas, murinas…) como a los distintos abordajes experimentales empleados. Para su sobreexpresión se han usado tanto métodos directos, como el uso de telómeros transcripcionalmente inducibles [214] o indirectos usando telómeros artificiales [215] o suplementando cultivos celulares con secuencias teloméricas de RNA artificiales [176]. En cambio, para su silenciamiento aunque se han usado silenciamientos directos mediante el uso de siRNAs o Gapmers [167,173,216], mayoritariamente se han usado métodos indirectos, basados en sobreexpresar o silenciar, proteínas capaces de modular los niveles de TERRA [163,165,174]. El mayor problema de estos sistemas de silenciamiento directo, es su limitada efectividad, la dificultad en su reproducción y que tienen un límite de silenciamiento de entre algunas horas y dos días. Mientras que el problema de los silenciamientos indirectos, es la imposibilidad de distinguir entre los fenotipos derivados de los cambios transcripcionales de TERRA del que deriva de las alteraciones en las proteínas que modulan los niveles de TERRA.

A consecuencia de la falta de modelos KO, los problemas en la cuantificación de TERRA y le heterogeneidad de los modelos usados, es que a TERRA se le han asignado funciones en principio contradictorias entre ellas. Un ejemplo claro es el caso de su papel en el mantenimiento de la longitud telomérica, donde se ha propuesto que puede ser tanto un modulado negativo de la actividad telomerasa [152,166,167], pero también se ha propuesto que al transcribirse a partir de telómeros cortos [168] sería capaz de reclutar la telomerasa a los mismos, o de la misma manera favorecer la formación de R-loops y de la elongación mediada por ALT. Un exceso de estas estructuras podrían tener el efecto contrario [163,169,170]. Otra función no del todo clara de TERRA es la de su papel en la estabilidad telomérica, mientras que algunos estudios ponen de manifiesto que el silenciamiento de TERRA incrementa el daño y la inestabilidad telomérica [167,172,173], otras sugieran que el incremento de la formación de R-loops en el telómero serían una fuente de inestabilidad y aumentarían el daño [174,175]. Es difícil explicar esta aparente contradicción con respecto a las funciones de TERRA, pero podría deberse por una parte a la forma en la que TERRA interacciona con el telómero. Algunos métodos usados para el estudio de TERRA estarían interfiriendo con las moléculas de TERRA que actuarían en *cis*, es decir, en su telómero de origen (por lo tanto, interaccionando mayoritariamente a través de la formación de R-loops). Mientras que otros métodos afectarían a los TERRA que actuarían en *trans* (es decir, en localizaciones distintas a su origen).

1.2. Los transcritos de los subtelómeros 20q presentan características propias de TERRA

Como ya hemos mencionado la principal causa por las que se utilizan un rango tan amplio de metodologías es la dificultad de generar modelos genéticos KO para TERRA, al desconocerse hasta ahora si estas moléculas se expresan de todos los telómeros o de algunos en concretos y por la falta de información sobre su origen genético. Por eso, en el Artículo 1 nos centramos en encontrar el origen transcripcional de TERRA en humanos.

En células humanas los primeros estudios, situaron la transcripción telomérica en subtelómeros cuyos promotores se basaban en islas CpG [157]. Posteriormente mediante el uso de células HeLa en las cuales se incrementó la expresión de TERRA al silenciar TRF2 se localizaron 18 posibles subtelómeros que transcribían para TERRA de dos tipos distintos aquellos donde la transcripción empezaba entre 5-10 Kb del telómero y aquellos donde empezaba a 1Kb o menos [160]. Esta localización se hizo a partir de un pull-down de RNA usando una sonda contra la repetición telomérica seguido de RNA-seq. Tras el alineamiento de estas secuencias, se consideraron regiones codificantes para TERRA todos aquellos subtelómeros enriquecidos en estas secuencias a menos de 10 Kb del telómero, pero este estudio no confirmaba si estos transcritos presentaban otras características propias de TERRA [160]. La falta de confirmación de características propias de TERRA de los trascritos propuestos como TERRA en este estudio, nos llevó a validar si alguno de estos *loci* presentaba características de TERRA, con la idea de posteriormente poder eliminarlos y generar modelos TERRA KO. Este tipo de abordaje experimental ya se ha desarrollado previamente en nuestro laboratorio en el caso de células de ratón donde en MEF e IPSCs se caracterizó que la transcripción telomérica se da mayoritariamente en los subtelómeros del cromosoma 18 y 9 [172] o en otros donde han observado que en el caso de ESC esta transcripción podría ocurrir a partir de la región PAR de los cromosomas sexuales [178].

Para reevaluar el RNA-seq previo usado para identificar TERRA [160], alineamos sus lecturas en dos *genome browsers*, el *browser* humano de USCS GRCh38/hg38 y un *browser* específico para subtelómeros [217] (ya que al ser estructuras altamente repetitivas, presenta muchos problemas de secuenciación en los *browsers* convencionales). Esto nos permitió comprobar que el 83,3% de los *loci* descritos en el estudio en cuyo RNA-seq nos estábamos basando [160] estaban en subtelómeros estructuralmente conservados. Los 10 Kb TERRA conservados presentaban tres elementos, por orden de cercanía al centrómero, pseudogenes pertenecientes a la familia WASH, pseudogenes pertenecientes a la familia DDX11L y una región repetitiva denominada TAR1 que se presenta justo antes del inicio del telómero, mientras que los 1 Kb TERRA solo presentaban la región TAR1. El hecho de que entre esos elementos se encontrara el gen codificante para proteínas WASH y que los pseudogenes de la familia DDX11L se transcriban en sentido telómero-centrómero, el opuesto al sentido transcripcional de TERRA, revelaba tenían características distintas a las de TERRA. Para validar si los transcritos derivados de estos locus presentaban características propias de TERRA como puede ser la presencia de una repetición telomérica UUAGGG o la heterogeneidad en el tamaño de sus moléculas,

aprovechamos el hecho de que estos subtelómeros estaban estructuralmente conservados y diseñamos sondas de RNA-FISH y Northern capaces de reconocer a la región TAR1 o a los diferentes pseudogenes de la familia DDX11L y WASH de todos estos subtelómeros. Sin embargo, no pudimos encontrar características propias de TERRA en ninguno de estos transcritos. Entre las posibles explicaciones para esto, se encuentran que estos transcritos sean artefactos derivados del proceso de pull-down, probablemente debido a que la sonda contra la repetición telomérica de TERRA es altamente repetitiva y podría hibridar con alguna de las secuencias de estos elementos o contra las propias ITS que se encuentran con frecuencia en el subtelómero. Otra posible opción, es que la cantidad de moléculas de TERRA transcritas a partir de estos subtelómeros sea tan pequeña que no lleve al límite de detección mínimo requerido para el RNA-FISH o el Northern.

Sin embargo, cuando evaluamos los transcritos provenientes de los subtelómeros no conservados (20q, Xp, 17p) pudimos encontrar colocalización entre la sonda contra la repetición telomérica de TERRA y la sonda contra la secuencia de RNA subtelomérica de los subtelómeros 20q y Xp. Esto pone de manifiesto la posibilidad de que estos subtelómeros fueran bona fide *TERRA loci*. Cabe destacar varios aspectos de estos *locus*, el primero es que los RNAs transcritos a partir del subtelómero 20q y Xp no muestran una homología de secuencia entre ellos. El segundo es que estudios posteriores a la publicación del Artículo 1 que compone esta tesis, han eliminado esta región del subtelómero 17p en células humanas HCT111 [156]. Estas células 17p-KO presentan también fenotipo telomérico, en concreto problemas a la hora de replicar el telómero. A pesar de ello, en este estudio no han probado que la eliminación del subtelómero 17p genere una caída de las TERRA totales, aunque si son capaces de rescatar el fenotipo al sobreexpresar TERRA de forma indirecta [156]. Esto apunta a que en diferentes tipos celulares diferentes *loci* podrían estar contribuyendo a la población de TERRA total. La principal característica del subtelómero 17p, es la presencia de una zona de unión de CTCF, característica que comparte con la región promotora Pr1. Esto lleva a pensar que como estaba descrito previamente [155] CTCF podría estar mediando de forma muy importante la expresión de TERRA o en qué subtelómeros se transcriben en cada célula. El último aspecto interesante de estos tres *loci*, es que la región del subtelómero Xp, es una de las regiones PAR de los cromosomas sexuales humanos. Como hemos mencionado antes, también se han evidenciado características propias de TERRA de los transcritos de las regiones PAR de los cromosomas sexuales de ratones [178], por lo que se podría pensar en una conservación evolutiva en la transcripción telomérica a partir de los cromosomas sexuales de mamífero, Sin embargo, no hemos podido encontrar una conservación estructural o evolutiva entre el *loci* 20q identificado en esta tesis y los otros *loci* identificados en ratón, los del subtelómero 18 y del subtelómero 9.

1.3. El subtelómero 20q es un *bona fide* TERRA *locus*

La evidencia más importante de que el subtelómero 20q es un *bona fide* TERRA *locus* es que, la eliminación del mismo en homocigosis en células humanas U2OS genera una caída notable en los niveles de TERRA totales. Elegimos las U2OS como modelo de estudio por dos

razones principales, la primera es que son las células humanas de cáncer donde hasta ahora se han reportado los mayores niveles de TERRA al elongar los telómeros vía ALT [161,162]. La otra es que al estar el ALT tan asociado a la expresión de TERRA, abre la posibilidad de que este último sea una posible diana terapéutica para estos tipos de tumores, por lo tanto, la caracterización del papel de TERRA en estas células resulta especialmente interesante. Los diferentes clones de células 20q-TERRA KO presentaban diferentes niveles de caída en los niveles de TERRA totales, siendo la caída mayor en el caso del clon C4. Esto es probablemente debido a la variedad intrínseca de las células que componen las líneas celulares de cáncer y a procesos de adaptación durante la expansión monoclonal. Aunque conseguimos general células U2OS 20q-TERRA KO, fallamos en hacer lo mismo en otras líneas celulares humanas telomerasa positivas, como pueden ser las HeLa o las HCT116. Esto se puede deber a que al poseer estas células unos niveles basales de TERRA muy bajos, los clones sin esta región no puedan tolerar la pérdida de TERRA y mueran. Otra posibilidad es la baja eficiencia del proceso de generación de clones homocigotos para grandes deleciones mediante CRISPR en el momento de realización de ese experimento. Esta falta de eficiencia ha quedado resuelta con el uso plásmido que codifica todos los elementos necesarios para la generación de la deleción. Esto nos ha permitido producir un mayor número de clones de U2OS 20q-TERRA KO. Este sistema mejorado, podría usarse en el futuro para generar diferentes líneas celulares sin esta región, para confirmar nuestros datos anteriores o ampliarlos a sistemas celulares positivos para la telomerasa.

Cabe destacar que, aunque los clones 20q-TERRA KO presentaban una bajada significativa para los TERRA totales, no observamos lo mismo en el caso del clon Xp-KO que generamos. Esto podría deberse a dos motivos. A que el locus Xp no estuviese transcribiendo para TERRA y a que las colocalizaciones que observamos fueran debidas a un artefacto de la sonda contra los subtelómeros de este transcrito que empleamos en este estudio. El hecho de que no fuéramos capaces de diseñar un gRNA para el sistema CRISPR completamente adyacente al telómero Xp, debido al carácter poco conservado y altamente repetitivo del mismo podría ser otra de las causas. Estas limitaciones nos impidieron delecionar completamente ese locus y quizás el promotor de TERRA quedó inalterado.

Estos modelos celulares representan la primera aproximación genética y los primeros KO para TERRA en cualquier organismo y, nos ha permitido ampliar ampliamente nuestro conocimiento de las implicaciones de este lncRNA en el mantenimiento del telómero y su epigenética. Este modelo presenta bastantes características ventajosas con respecto a los usados hasta ahora, una de las más importantes es la estabilidad del mismo, mientras que otros modelos de sobreexpresión o silenciamiento como pueden ser los Gapmers, requieren de modificar los niveles de TERRA *de novo* cada vez que se realiza un experimento. Los niveles de TERRA se mantienen establemente bajos en nuestras células, facilitando el trabajo experimental y permitiendo el estudio del efecto de la eliminación de TERRA a largo plazo. Otra ventaja importante es que es un silenciamiento directo de TERRA, no mediado por la sobreexpresión o eliminación de moduladores de TERRA. Esto permite ver de forma clara la implicación de TERRA en los distintos procesos, sin los efectos colaterales de alterar otras proteínas. Este modelo

también carece de los *off-targets* derivados de silenciar la repetición telomérica de TERRA UUAGGG mediante siRNAs o Gapmers, ya que esta secuencia se haya repartida por todo el genoma y probablemente presente en otros transcritos [32–34]. Una idea importante derivada de este modelo, es que es posible obtener células KO para TERRA. Esto abre la posibilidad de generar animales KO para TERRA, en concreto modelos murinos, donde ya se han localizado tres loci diferentes que se podrían delecionar [172,178]. Estos animales serían un modelo idóneo para estudiar el papel fisiológico de la transcripción telomérica y solventaría algunos de los problemas derivados del uso de modelos celulares y en concreto los posibles de *off-targets* causados por el sistema CRISPR, al poder eliminarlos mediante cruces.

A pesar de todas estas ventajas, el modelo tiene limitaciones, algunas intrínsecas a la técnica y otras de carácter biológicas. Una de las primeras es los posibles *off-targets* derivados del sistema CRISPR. Para solucionarlos hemos tomado tres medidas principales a lo largo de este trabajo: la primera de ella es el uso de sistemas CRISPR codificados en vectores episomales, no integrativos. Esto minimiza el tiempo en que la enzima Cas9 se encuentra activa en la célula y por lo tanto reduce los posibles eventos de *off-targets*. La segunda es la selección de gRNAs que presentan la mínima homología posible con otras regiones, a través de su diseño en herramientas informáticas especialmente diseñadas para ello [218]. Y la tercera fue el análisis de posibles *off-targets* en nuestras líneas celulares mediante dos aproximaciones, el uso del *surveyor nuclease assay*, y del análisis por TIDE [219]. Además, durante estos años la generación de mutantes mediante CRISPR se ha vuelto una técnica aceptada y de uso general en diferentes laboratorios. La otra desventaja técnica, es que el tiempo de generación de líneas celulares a partir de células individualizadas, hace imposible el uso de esta metodología para estudiar eventos que ocurren a tiempos cortos. A esto se le suma que las poblaciones celulares, pueden cambiar a lo largo de los pases y acumular mutaciones y cambios epigenéticos. Para intentar reducir el efecto de la expansión monoclonal en nuestros resultados, hemos comparado nuestras células KO con el pool celular original y mantenido en cultivo el mismo tiempo que tomó la expansión monoclonal y con clones WT expandido a la vez que los KO. Una de las mayores limitaciones biológicas es que al no estar el locus *20q* conservado entre especias, imposibilita su uso más allá de células humanas. También, diferentes células humanas podrían presentar diferentes *locus* mayoritarios para TERRA, dificultando la generación de células TERRA-KO, como ocurre en el caso de levadura donde se descrito que solo un telómero podría estar trascribiéndose cada vez [168]. Para solucionar estos dos problemas se necesitaría de otras formas para localizar posibles *loci* de TERRA, ya que el uso de pull-down contra TERRA seguido de RNA-seq genera un gran número de falsos positivos e incluso sus secuencias alinean con regiones intracromosomales, como hemos comprobado a lo largo de este trabajo y en trabajos previos del laboratorio [172]. Esta localización se podría mejorar mediante el uso de CHIRT-seq, que como se puede apreciar en el artículo 3 de esta tesis y en trabajo de otros laboratorios [167], es capaz de detectar las uniones en *cis* de TERRA e identificar *loci* previamente reportados en ratón como es el caso del cromosoma 18 [172] o de la región PAR de los cromosomas sexuales [178].

1.4. TERRA es importante para el mantenimiento de la estabilidad y de la longitud telomérica

Las líneas celulares 20q-TERRA KO nos han permitido estudiar de forma directa el papel de TERRA en el mantenimiento del telómero. La longitud telomérica es una parte clave para que el telómero pueda cumplir correctamente con su papel celular, especialmente en células altamente proliferativas como las tumorales. Tras eliminar TERRA observamos una importante reducción de longitud telomérica acompañada de una acumulación de telómeros cortos. A su vez vimos una caída de las T-SCE en estas células, indicando que el efecto podría ser debido a una pérdida del ALT. Estos datos están en consonancia de varios trabajos publicados hasta ahora, que han descrito una gran importancia de TERRA para que ocurra la recombinación telomérica y el ALT [163,169,170,215]. Al estar usando células que elongan los telómeros exclusivamente vía ALT no podemos extrapolar que el efecto de la eliminación de TERRA en la longitud telomérica sea algo común para todos los tipos celulares, ya que la elongación por telomerasa es un proceso completamente independiente al ALT. Para solventar esto, como ya hemos mencionado sería necesaria la generación de células telomerasa positiva humanas 20q-TERRA KO. Este método, podría resolver la actual duda del papel de TERRA como regulador de la telomerasa, ya que algunos estudios evidencian que TERRA podría funcionar como un modulador negativo del mismo [152,166,167], mientras que en levaduras se ha propuesto que TERRA podría reclutar a la telomerasa a telómeros cortos favoreciendo su elongación [168]. Finalmente, con las evidencias experimentales que tenemos hasta ahora, no podemos excluir la posibilidad de que esto estuviera producido por problemas replicativos en el telómero, ya que está descrito que TERRA podría ser importante en la replicación telomérica y esta no se ha estudiado en nuestro modelo [156,171]

El segundo fenotipo telomérico importante que hemos confirmado con nuestro modelo de células 20q-TERRA KO es un incremento en el daño en el DNA, tanto en el telomérico como en el total. Este incremento del daño está acompañado de un incremento de la fusión entre cromosomas y de la inestabilidad génica. Este último fenotipo, solo ha aparecido en el caso del clon C4, no en los otros y esto es debido probablemente a que este es el que presenta el mayor silenciamiento de TERRA. La pérdida de la estabilidad telomérica y cromosómica está a favor de varios estudios, que ven un fenotipo similar en células humanas y de ratón al silenciar tanto TERRA de *locus* específicos como los TERRA totales mediante siRNAs [173] o Gapmers [167,216]. Sin embargo, contradice otra serie de estudios, donde han observado que un incremento de la transcripción telomérica favorece la inestabilidad génica, sobre todo por la formación de R-loops [174,175]. La aparente contradicción en el papel de TERRA en la estabilidad telomérica y cromosómica, señalando algunos estudios como necesario para la protección telomérica y otros como una fuente de inestabilidad al formar R-loops, podría estar explicada por los diferentes abordajes experimentales usados. Los métodos directos de silenciar TERRA pueden inferferir con moléculas de TERRA que interaccionaran con el telómero tanto en *cis* como en *trans*, por lo que estas uniones podrían ser directas a través de la formación de R-loops o G-cuadruples [174,175,220], pero también indirectas mediante la interacción con las shelterinas o componentes de

la heterocromatina telomérica [173]. Este tipo de experimentos ha puesto de manifiesto que TERRA es importante para el mantenimiento telomérico y que el silenciamiento de los TERRA totales o específicos para un locus, genera un incremento del daño general y del daño en todos los telómeros [167,172,216] y la pérdida de longitud telomérica en ESCs [167]. En cambio, los estudios llevados a cabo para entender el papel de TERRA a través del establecimiento de R-loops, están basados fundamentalmente en el silenciamiento o sobreexpresión de shelterinas como TRF1 o TRF2 [174] o de la RNAseH1 [163], por lo que solo afectaría a las moléculas de TERRA formando R-loops y mayoritariamente unidas al telómero en cis. Estos datos sugieren la posible existencia de dos poblaciones distintas de moléculas de TERRA. Algunas moléculas de TERRA estarían interaccionando directamente con su telómero de origen formando R-loop, mientras que otras moléculas de TERRA se unirían en *trans* a telómeros diferentes de aquellos de los que se trascribieron u otras regiones del genoma. Al formar RNA-loops TERRA desestabilizaría el telómero, hasta que estos se resolvieran, en algunos casos incluso vía ALT. En cambio, las moléculas de TERRA funcionando en *trans*, podrían estar reclutando diferentes proteínas a los telómeros favoreciendo su reparación o estabilidad. La gran movilidad y dinámica de las moléculas de TERRA ya se ha observado de forma experimental a través del seguimiento de las mismas en células vivas, pudiendo pasar éstas de estar en otras partes del núcleo, a unirse con distintos telómeros, o incluso a formar agregados entre ellas [216].

Otra parte importante de estos datos es que los hemos podido validar en un segundo set de clones 20q-TERRA KO, los generados para el Artículo 2. En estos clones hemos podido ver al igual que en los generado en el Artículo 1, una caída de la longitud telomérica, así como un incremento del daño total y telomérico (datos no publicados).

1.5. TERRA facilita el establecimiento de la heterocromatina telomérica a través del reclutamiento del complejo PRC2

El último aspecto que abordamos usando las células 20q-TERRA KO fue el establecimiento de la heterocromatina telomérica. En el Artículo 2 describimos que TERRA es importante para el establecimiento de las marcas de la heterocromatina telomérica (H3K9me3, H4K20me3 y HP1) y para la marca del complejo PRC2, H3K27me3. Además, demostramos que el papel de TERRA en la cromatina telomérica está mediado a través del reclutamiento del complejo PRC2.

Diferentes estudios han puesto de manifiesto que el establecimiento de la cromatina telomérica tiene una alta importancia en el comportamiento del telómero. Un telómero altamente heterocromatinizado favorece el TPE y la protección telomérica [76,81,109,112,132–135], mientras que la pérdida de esta heterocromatina favorecería la acción de la telomerasa y la recombinación telomérica (evento asociado normalmente con una activación del ALT) [79–81,112,130]. A pesar de que la perdida de la heterocromatina telomérica resulta en un incremento de la recombinación telomérica, el papel de la heterocromatina telomérica en el sistema ALT no está claro, ya que también se ha descrito que la unión de HP1 al telómero podría ser importante para el ALT, además de que las células ALT humanas podrían tener telómeros más heterocromatinizados que

las células telomerasa positivas [93,131]. TERRA ya ha sido anteriormente implicado en el proceso de heterocromatinización del telómero ya que es capaz de interacción con H3K9me3, HP1 y con Suv39h1 [173,160]. El silenciamiento de TERRA causa la pérdida de heterocromatina telomérica [173], mientras que su sobreexpresión indirecta al silenciar TRF2 coincide con un incremento de H3K9me3 en el telómero [160]. Nuestros datos, que revelan una pérdida de las marcas asociadas a la heterocromatina telomérica (H3K9me3, H4K20me3 y HP1) al reducir los niveles de TERRA, estarían respaldadas por estos estudios y reforzarían la idea de que TERRA es importante para la heterocromatina telomérica.

Más allá de confirmar el papel de TERRA como un modulador clave para la deposición de heterocromatina telomérica, este trabajo aporta un conocimiento más profundo del proceso de heterocromatinización telomérica. Al eliminar TERRA, nuestras células también presentan la caída de la marca de heterocromatina facultativa H3K27me3 en los telómeros. Esta marca solo había sido observada previamente en los telómeros de eucariotas inferiores como las algas unicelulares [223] y en los subtelómeros del hongo filamentoso *Neurospora crassa* [224]. Dado que esta marca es establecida en exclusiva por el complejo PRC2 y este complejo está implicado en establecer regiones heterocromáticas junto a distintos lncRNAs incluyendo la compactación de grandes estructuras de DNA junto a XIST [225–230], dio pie a pensar que el complejo PRC2 podría estar implicado en la función de TERRA sobre la epigenética telomérica. Así. confirmamos que TERRA es capaz de interaccionar con las subunidades del complejo PRC2, SUZ12 y EZH2. Esta interacción también ha sido valida *in vitro* e *in vivo* en células humanas y de ratón, siendo los RNAs capaces de formar G-cuádruples como TERRA los que se unen con más afinidad a este complejo [167,231]. Usando inmunofluorescencia confocal demostramos que la interacción (hasta ahora no descrita) del complejo PRC2 con el telómero era dependiente de TERRA. La capacidad de TERRA de reclutar a PRC2 está en parte reforzado por un estudio en el que se ha descrito que TERRA es capaz de unirse por todo el genoma y que su unión está enriquecida en sitios donde la marca H3K27me3 se encuentra presente [167]. Finalmente, pudimos confirmar la importancia de PRC2 en el telómero, ya que al silenciar mediante shRNAs distintas subunidades del complejo PRC2, se perdía la heterocromatina telomérica en células U2OS.

Sorprendentemente, en las células 20q-TERRA KO que presentan menos marcas de heterocromatina en el telómero, hemos observado una pérdida de las T-SCE. En cambio, en otros modelos celulares o animales donde se reduce esta heterocromatina telomérica, se produce un aumento de las T-SCE [79–81,112,130]. Esto revela que TERRA podría ser la clave detrás del aumento de la recombinación telomérica en estos modelos, ya que la descompactación telomérica también se ha asociado con un incremento de la expresión de TERRA [152,157–159]. Por eso, al estar eliminando TERRA, nosotros estaríamos observando una caída de las T-SCE a pesar de la caída en la heterocromatina telomérica. Además de en la longitud telomérica, otro efecto donde podríamos estar influyendo la pérdida de la heterocromatina que vemos en nuestras células podría ser en el incremento del daño, explicando esto en parte porque observamos un aumento del daño telomérico en nuestro modelo 20q-TERRA KO.

Estos datos en su conjunto, nos llevan a proponer un modelo por el cual TERRA sería responsable de establecer la heterocromatina telomérica al reclutar al complejo PRC2 al telómero. Sin embargo, el reclutamiento del complejo PRC2, podría no ser el único papel que estuviese teniendo TERRA, sino que de forma complementaria podría estar reclutando otros componentes necesarios para este proceso como HP1 y Suv39h1, tal y como han descrito otros estudios [173,160]. Establecer este modelo como algo general para todas las células humanas es complicado, sobre todo para el caso de las células telomerasa positiva. Esto ocurre por tres motivos principales, el primero es que las células ALT y las células telomerasa positivas humanas tienen una epigenética telomérica muy distinta, de hecho parece que los telómeros de las células ALT están más heterocromatinizados [93]. Este carácter más heterocromatinizado, podría deberse a la presencia del mecanismo descrito en nuestro trabajo, ya que las células ALT presentan niveles mucho más altos de TERRA que las células tumorales telomerasa positivas [161,162]. El segundo motivo es que las U2OS, como muchas otras líneas celulares tipo ALT carecen de ATRX y ATRX es necesario para la correcta unión del complejo PRC2 por todo el genoma [232]. Al carecer estas células de ATRX podría haber mucho más complejo PRC2 libre en la célula. A esto se le sumaría el tercer motivo, los niveles de TERRA no están regulados través del ciclo celular en células ALT al contrario que en células telomerasa positivas donde los niveles de TERRA alcanzan su máxima abundancia en la fase G1 y decreciendo a lo largo de la fase S [154]. Esta falta de regulación a través del ciclo celular, lleva a que TERRA esté constitutivamente expresado en las U2OS [233]. Esto favorecería una interacción más constante de moléculas de TERRA con el telómero que junto a una mayor disponibilidad del complejo PRC2, podría hacer este mecanismo exclusivo de células ALT. Por estos motivos sería especialmente interesante confirmar este mecanismo en células telomerasa positivas, ya fuese generando más modelos TERRA KO o silenciando directamente el complejo PRC2 en las mismas.

1.6. Posibles aplicaciones de las células 20q-TERRA KO para entender el papel de TERRA en cáncer

Los modelos TERRA-KO generados en esta tesis, nos podían permitir ampliar nuestro conocimiento de los mecanismos moleculares en los que está implicado TERRA. Por ejemplo, validando como puede regular el sistema ALT. Con respecto al daño en el DNA, sería especialmente interesante estudiar a partir de que eventos se producen, si este es daño por *uncapping* telomérico, daño replicativo, por rotura de doble cadena, etc. Esto nos permitiría pensar en posibles estrategias terapéuticas que combinaran eliminar TERRA y los mecanismos que reparen el posible daño que causa, para tratar tumores tipo ALT. Dado que observamos un gran incremento del daño total, también sería interesante estudiar si este es debido solamente al daño telomérico o podría responder a funciones extratelomércas de TERRA, que actuara junto a algunos sistemas de reparación del daño en el DNA. Estas posibles funciones extratelomércas estarían conceptualmente apoyadas por el hecho de que TERRA se capaz de unirse por todo el genoma [167] y a que los niveles de TERRA aumenten ante distintos estreses celulares o ante el incremento en el DNA o *uncapping* telomérico ([152,160,221,222]).

2. Estudio del papel de la proteína telomérica TRF1 en pluripotencia

2.1. TRF1 es esencial para la reprogramación y el mantenimiento de la pluripotencia

Para el correcto funcionamiento del telómero, al mismo se tienen que unir un complejo de proteínas llamadas shelterinas. Este complejo es esencial para la protección del telómero, evitando que sea reconocido como daño de doble cadena y con ello impidiendo la activación del DDR [16]. Entre las proteínas que conforman el complejo shelterina se encuentra TRF1. TRF1, además de estar implicada en evitar la activación de DDR en el telómero también, tiene otras funciones teloméricas, como regular la longitud telomérica y su correcta replicación [26,29,30]. Más allá de estas funciones, diferentes estudios han puesto de manifiesto que TRF1 podría estar también implicada en pluripotencia. OCT4 se une al promotor de TRF1 [209] y TRF1 está sobreexpresado en ESCs, iPSCs y durante la reprogramación [206,209,211,212]. Otras evidencias más directas, son que TRF1 es importante para la reprogramación *in vitro* e *in vivo* [206,209] y además los ratones KO para TRF1 presentan letalidad embrionaria en el estado de blastocito, sin que esta letalidad esté asociada a un fenotipo telomérico [210], sugiriendo que estas funciones serían independientes a su papel en la protección del telómero.

En el Artículo 3 observamos que el silenciamiento de TRF1 en iPSCs en estado naïve, producía el cambio de expresión de cientos de genes asociados a pluripotencia y diferenciación. Este cambio transcripcional estaba asociado a un incremento por todo el genoma y en genes cuya expresión era dependiente de TRF1 de la unión del complejo PRC2 y de depósito de su marca H3K27me3. El silenciamiento de TRF1, también producía un incremento de la expresión de TERRA y a su vez TERRA se unía en las cercanías de los genes cuya expresión era dependiente de TRF1. Esto, nos ha llevado a proponer un modelo por el cual TRF1 es esencial para la pluripotencia al regular la expresión de cientos de genes de una forma dependiente de TERRA. Esto lo haría al modular los niveles de TERRA y este a su vez favorecería el reclutamiento del complejo PRC2 a distintos genes asociados con pluripotencia y diferenciación, alterando su expresión.

2.2. La eliminación de TRF1 en IPSCs desencadena cambios drásticos en la expresión génica y en la epigenética celular

Para tratar de elucidar el papel de TRF1 en pluripotencia usamos iPSCs. El silenciamiento de TRF1 lo hicimos mediante el uso de un shRNA contra TRF1 y para poder distinguir el efecto directo de TRF1 del efectuado por la señalización de daño en el DNA, se realizó en iPSCs P53 KO. En todos los experimentos las células se recogieron en el tiempo mínimo necesario para poder seleccionar las infectadas. Debido a que las iPSCs P53 KO tienden a diferenciarse espontáneamente en cultivo [234–237], trabajamos con iPSCs en estado naïve, lo que las hace más estables [235–237].

Nuestra primera aproximación para profundizar en el mecanismo por el cual TRF1 es esencial para el establecimiento de la pluripotencia, fue comprobar si el silenciamiento de TRF1

causaba cambios transcripcionales. Para ello realizamos RNA-seq del transcriptoma completo de nuestras iPSCs infectadas con un vector *scramble* o con un shRNA contra TRF1. Al comparar los resultados de las células con silenciamiento de TRF1 con respecto al control, observamos un cambio significativo en la expresión de cientos de genes, sobreexpresandose 483 genes y bajando la expresión de 328. Por GSEA y por Enrichr, validamos que estos genes estaban fundamentalmente asociados a pluripontecia, a diferenciación y varias rutas de señalización asociadas a este proceso. Confirmamos que estos efectos no eran debidos a un aumento en el daño en el DNA, ya que por inmunofluorescencia no pudimos observar un aumento del daño en el DNA total o telomérico, indicando que el proceso de señalización por *uncapping* telomérico no se había iniciado.

Debido a que el 64% de los genes que se sobreexpresaban al silenciar TRF1 presentan unión del complejo PRC2 en sus promotores, pensamos que estos cambios podrían estar asociados a una alteración de la unión de este complejo. Por ChIP-seq observamos un aumento de la unión de PRC2 por todo el genoma al eliminar TRF1 y también un incremento de su marca, H3K27me3. Este incremento de la unión de PRC2 también ocurría en los promotores de los genes cuya expresión cambiaba al silenciar TRF1, sugiriendo ésto que su cambio de expresión se podría deber al complejo PRC2. Cambios en los niveles de H3K27me3, ya se habían observado previamente en relación al telómero. ESCs murinas con telómeros cortos muestran un ligero aumento de los niveles totales de H3K27me3 y además fallan a la hora de diferenciarse adecuadamente [203]. Esto coincidiría con nuestros datos y sugeriría un modelo donde ciertas alteraciones teloméricas podrían alterar los niveles de H3K27me3 modulando la pluripotencia. Esto es especialmente interesante ya que nuestros datos sugieren que estas alteraciones son debidas a un aumento de la expresión de TERRA. Estudios previos han demostrado que los telómeros cortos también promueven la expresión de TERRA [159], lo que también podría explicar el incremento de H3K27me3 en ESCs con telómeros cortos.

A pesar de que el complejo PRC2 y su marca H3K27me3 están principalmente asociados a la represión transcripcional, tras eliminar TRF1, observamos un conjunto de genes cuya expresión aumenta a la vez que aumenta tanto la unión de PRC2 como el depósito de su marca. Esta aparente contradicción puede explicase de dos formas distintas. La primera forma es que varios estudios han reportado que distintas subunidades de PRC2 puede tener funciones adicionales, como la activación transcripcional. Esto ocurre en algunos casos de cáncer de próstata resistentes a castración o en el cáncer de mama, donde cumple esta función de activación transcripcional a través de la interacción con distintos factores, como el receptor de estrógenos α, la β catenina, MYC o CCND1 [238,239]. La segunda es que aproximadamente el 70% de los genes que se sobreexpresan y tienen un aumento de la deposición de H3K27me3 en sus promotores son genes bivalentes. Estos son genes cuyos promotores presentan dos tipos de marcas de histonas, una marca activadora, la H3K4me3 y la marca represiva de PRC2, H3K27me3 [240–245]. Estas dos marcas se encuentran en genes de diferenciación celular y son necesarias para la rápida expresión de los mismos cuando estas células se tienen que diferenciar y su posterior silenciamiento cuando estas ya se han diferenciado [240,241]. Además, el

establecimiento de los promotores bivalentes es una característica importante del paso del estado naïve de las ESCs al estado primed [246]. En relación a esto, en nuestro modelo de iPSCs en las que silenciamos TRF1 también observamos un incremento de H3K27me3 en los promotores de estos genes bivalente, reforzando la idea de que al eliminar TRF1 se pierde la pluripotencia y el estado naïve.

Los cambios de expresión en genes asociados a pluripotencia y diferenciación, explicarían la importancia de TRF1 en pluripotencia. Sin embargo, otras cuestiones se podrían plantear con respecto a nuestro modelo. que podrían ser estudiadas como más profundidad. A pesar de que la mayoría de los genes cuya expresión cae estarían asociados a pluripotencia y los que se sobreexpresan a diferenciación no tenemos datos del destino celular que tendrían nuestras iPSCs sin TRF1. Si bien sería lógico pensar que estas tenderían a diferenciarse espontáneamente en cultivo, la pluripotencia es la capacidad de una célula de diferenciarse en cualquiera de las tres capas germinales, por lo que una pérdida de la pluripotencia también podría estar asociada a una incapacidad de estas células de diferenciarse correctamente. Para confirmar esto, sería interesante estudiar *in vitro*, a través de ensayos de diferenciación qué ocurriría en estas células, o *in vivo*, haciendo uso de experimentos de complementación de embriones o experimentos de formación de teratomas. Otra carencia de nuestro modelo, es que al tener las iPSCs un porcentaje de infección bajo, tenemos que seleccionarlas, por lo que tras la primera infección transcurren 4 días antes de poder observar los cambios en expresión génica. Esto nos impide discernir los eventos iniciales o solamente debidos a cambios transcripcionales tras silenciar TRF1. Para mejorar esto, podríamos recurrir a otros modelos celulares, en concreto podríamos aprovechar el modelo TRF1 KO condicional ($TRF1^{flox/flox}$) generado anteriormente en nuestro laboratorio [29]. Usando iPSCS o ESCs generadas a partir de este modelo cruzado con un animal transgénico UBC-Cre-ERT2, podríamos directamente delecionar TRF1 *in vitro* al añadir tamoxifeno a estas células sin necesidad de selección. Esto se podría combinar además con técnicas de secuenciación de una sola célula, lo que nos permitiría entender los cambios iniciales en expresión génica al eliminar TRF1. Por último, también sería interesante comprobar si estos mismos cambios ocurren en iPSC en estado primed o están limitados al estado naïve.

Tras validar que la importancia de TRF1 en la adquisición y mantenimiento de la pluripotencia es debido fundamentalmente a su control sobre la expresión de genes implicados en pluripotencia y diferenciación a causa de un cambio en la estructura de la cromatina por todo el genoma, específicamente de la interacción del complejo PRC2 y de la deposición de H3K27me3, nos quedaba entender como una proteína localizada en el telómero podría estar causando alteraciones por todo el genoma.

2.3. TRF1 es importante para el mantenimiento de la pluripotencia a través de la regulación de TERRA

La primera hipótesis de como TRF1 podría estar detrás de estos grandes cambios epigenéticos y de expresión, fue la interacción directa de TRF1 con el complejo PRC2. Esta

hipótesis fue descartada al no conseguir validar esta interacción por espectrometría de masas. Nuestra segunda hipótesis fue que al igual que ocurre en *Arabidopsis thaliana*, donde las proteínas homologas a TRF1, las TRBs son capaces de unirse directamente a diferentes promotores y reclutar al complejo PRC2 [247,248], TRF1 podría estar haciendo algo equivalente. Por ChIP-seq observamos la unión de TRF1 a algunas zonas extratelomércas, especialmente a ITS, pero las uniones no eran lo bastante abundantes para explicar un cambio epigenético y transcripcional tan general como el que observamos al silenciar TRF1. Por último, pensamos que el mecanismo podría ocurrir vía TERRA. Esto se fundamenta en estudios previos que han descrito que TERRA es capaz de interaccionar con TRF1 [167,173] y al Artículo 2 que compone esta tesis, donde describimos que TERRA es capaz de reclutar al complejo PRC2 al telómero, estableciendo de esta manera la cromatina telomérica. De hecho, cuando analizamos los niveles de TERRA por RNA-FISH, observamos un incremento en la expresión de TERRA al silenciar TRF1. Este incremento en la expresión de TERRA al eliminar TRF1 también se ha descrito en estudios previos [160,249,250]. Reforzando la idea de que nuestro fenotipo estaría mediado por TERRA, observamos un incremento de la unión de TERRA a genes implicados en pluripotencia y diferenciación, donde aumenta la unión de PRC2 y cuya expresión era dependiente de TRF1. Todos estos datos nos sugieren un modelo, a través del cual TRF1 controlaría los niveles de TERRA, y este a su vez se uniría a genes implicados en pluripotencia y diferenciación reclutando al complejo PRC2 y de esa manera regulando la expresión de los mismos.

A pesar de las pruebas experimentales que avalan este modelo, este estudio tiene algunas limitaciones que podrían ser abordadas de forma experimental. La primera de ellas es que no tiene en cuenta otras posibles formas por la que TERRA podría estar regulando la transcripción. Estudios previos ya han descrito que TERRA está implicado en la regulación transcripcional, pero el mecanismo que proponen es diferente al nuestro. El primero de esos estudios, propone que TERRA podría regular la transcripción al unirse a promotores capaces de formar G-cuádruplex, provocando la formación de estas estructuras y de esta manera alterando la transcripción [176]. El segundo estudio atribuye este papel de TERRA a la competición directa con ATRX, uniéndose TERRA a las mismas zonas que ATRX y por tanto desplazando a ATRX [167]. Estos dos mecanismos podrían ser complementarios al que proponemos y explicar parte de los cambios transcripcionales asociados al silenciamiento de TRF1. Por último, TRF1 podría estar alterando la expresión génica y la unión de PRC2 de otras formas, además de a través de TERRA. Por ejemplo, podría estar regulando la estructura 3D del genoma o la interacción del telómero con la envuelta nuclear (*tethering* telomérico) ya que TRF1 es importante para el correcto *tethering* durante la meiosis de espermatozoides [253].

Nuestro modelo por el cual TRF1 controlaría la expresión génica al regular los niveles de TERRA, podría aplicarse más allá del papel de TRF1 en pluripotencia y reprogramación, de esta forma, también sugiere un nuevo mecanismo, por el cual el estado del telómero podría decidir el destino de la célula al controlar la expresión de TERRA y éste a su vez la expresión de cientos de genes. Esto podría ocurrir al eliminar TRF1, pero también por otros motivos que afectan al telómero y que afectan la expresión de TERRA, como el acortamiento, el *uncapping* o el daño

telomérico [152,160,203,221,222]. El mecanismo que hemos descrito podría ser no solo importante para pluripotencia, sino que podría ocurrir en otros procesos celulares, como en cáncer. Nuestro laboratorio ha descrito que TRF1 es una diana terapéutica eficaz en el caso del cáncer pulmonar y el glioblastoma [251,252]. En el caso del glioblastoma, además, TRF1 también es esencial para las células madre tumorales, y su eliminación de TRF1 provoca la disminución del *stemmnes* de las mismas. Esta pérdida, también podría estar mediada a través de la regulación de TERRA y sería interesante confirmar este mecanismo en estos modelos y en otros como podrían ser las células madre de tejido adulto.

Conclusiones

Conclusiones

Estudio del papel de los Telomeric repeat-containing RNA (TERRA) en la biología del telómero:

-Un 83,3% de los *locus* previamente identificados como orígenes transcripcionales de TERRA están en subtelómeros estructuralmente conservados y sus transcritos no presentan características propias de TERRA.

-Tanto los RNAs transcritos a partir del subtelómero 20q como del Xp presentan características propias de TERRA. Además, la eliminación del locus del subtelómero 20q genera una drástica reducción en las moléculas de TERRA totales, confirmando que el subtelómero 20q es un *bona fide* TERRA *loci*.

-La reducción de los niveles de TERRA KO produce un incremento en el daño en el DNA total y en el telómero, y de una reducción en la longitud telomérica, lo que evidencia la importancia de TERRA en el mantenimiento telomérico.

-La reducción de los niveles de TERRA ocasiona una pérdida global de las marcas de heterocromatina telomérica, así como de la marca H3K27me3, por lo que TERRA es importante en el establecimiento de la heterocromatina telomérica.

-TERRA interacciona con el complejo PRC2 y favorece su interacción con el telómero. Esto junto a la perdida de marcas de heterocromatina telomérica tras silenciar PRC2, indica que la función de TERRA en el establecimiento de la cromatina telomérica está mediada por el complejo PRC2.

Estudio del papel de la proteína telomérica TRF1 en pluripotencia.

-TRF1 está implicado en la regulación de la expresión génica, sobre todo de genes asociados a pluripotencia y diferenciación.

-TRF1 modula la unión del complejo PRC2 y del depósito de H3K27me3 por todo el genoma, incluyendo los promotores de los genes cuya expresión cambia al silenciar TRF1.

-TERRA se une a los promotores de parte de los genes cuya expresión está regulada por los niveles de TRF1.

-TRF1 regula el mantenimiento de la pluripotencia a través del control de la expresión de TERRA, que a su vez regula la expresión de genes asociados a pluripotencia y diferenciación.

Bibliografía

Bibliografía

1. Muller, J. H. The remaking of chromosomes. *Collect. Net* (1938).
2. McClintock, B. The Stability of Broken Ends of Chromosomes in Zea Mays. *Genetics* (1941).
3. Blackburn, E. H. & Gall, J. G. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J. Mol. Biol.* (1978).
4. Blackburn, E. H. Telomeres: no end in sight. *Cell*. (1994).
5. Walter, M. F. *et al.* DNA organization and polymorphism of a wild-type Drosophila telomere region. *Chromosoma* (1995).
6. Meyne, J., Ratliff, R. L. & Moyzis, R. K. Conservation of the human telomere sequence (TTAGGG)(n) among vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* (1989).
7. Klobutcher, L. A., Swanton, M. T., Donini, P. & Prescott, D. M. All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proc. Natl. Acad. Sci. U. S. A.* (1981).
8. Griffith, J. D. *et al.* Mammalian telomeres end in a large duplex loop. *Cell* (1999).
9. Flores, I. *et al.* The longest telomeres: A general signature of adult stem cell compartments. *Genes Dev.* (2008).
10. Marion, R. M. *et al.* Telomeres Acquire Embryonic Stem Cell Characteristics in Induced Pluripotent Stem Cells. *Cell Stem Cell* (2009).
11. de Lange, T. *et al.* Structure and variability of human chromosome ends. *Mol. Cell. Biol.* (1990).
12. Zijlmans, J. M. J. M. *et al.* Telomeres in the mouse have large inter-chromosomal variations in the number of t2ag3 repeats. *Proc. Natl. Acad. Sci. U. S. A.* (1997).
13. Hemann, M. T. Wild-derived inbred mouse strains have short telomeres. *Nucleic Acids Res.* (2000).
14. Blasco, M. A. *et al.* Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* (1997).
15. Lee, H. W. *et al.* Essential role of mouse telomerase in highly proliferative organs. *Nature* (1998).
16. De Lange, T. Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes and Development* (2005).
17. Palm, W. & de Lange, T. How Shelterin Protects Mammalian Telomeres. *Annu. Rev. Genet.* (2008).
18. De Lange, T. Protection of mammalian telomeres. *Oncogene* (2002).
19. Liu, D., O'Connor, M. S., Qin, J. & Songyang, Z. Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins. *J. Biol. Chem.* (2004).
20. Bianchi, A., Smith, S., Chong, L., Elias, P. & De Lange, T. TRF1 is a dimer and bends telomeric DNA. *EMBO J.* (1997).
21. Broccoli, D., Smogorzewska, A., Chong, L. & de Lange, T. Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nature Genetics* (1997).
22. Liu, D. *et al.* PTOP interacts with POT1 and regulates its localization to telomeres. *Nat. Cell Biol.* (2004).
23. Chen, Y. *et al.* A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science* (2008).
24. Kim, S. H. *et al.* TIN2 mediates functions of TRF2 at human telomeres. *J. Biol. Chem.* (2004).
25. Ye, J. Z. S. *et al.* TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *J. Biol. Chem.* (2004).
26. Smogorzewska, A. *et al.* Control of Human Telomere Length by TRF1 and TRF2. *Mol. Cell. Biol.* (2000).
27. Loayza, D. & De Lange, T. POT1 as a terminal transducer of TRF1 telomere length control. *Nature* (2003).
28. Tejera, A. M. *et al.* TPP1 is required for TERT recruitment, telomere elongation during nuclear reprogramming, and normal skin development in mice. *Dev. Cell* (2010).
29. Martínez, P. *et al.* Increased telomere fragility and fusions resulting from TRF1 deficiency lead to degenerative pathologies and increased cancer in mice. *Genes Dev.* (2009).
30. Sfeir, A. *et al.* Mammalian Telomeres Resemble Fragile Sites and Require TRF1 for Efficient Replication. *Cell* (2009).

31. Baumann, P. & Cech, T. R. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* (2001).
32. Meyne, J. *et al.* Distribution of non-telomeric sites of the (TTAGGG)n telomeric sequence in vertebrate chromosomes. *Chromosoma* (1990).
33. Nergadze, S. G., Santagostino, M. A., Salzano, A., Mondello, C. & Giulotto, E. Contribution of telomerase RNA retrotranscription to DNA double-strand break repair during mammalian genome evolution. *Genome Biol.* (2007).
34. Ruiz-Herrera, A., Nergadze, S. G., Santagostino, M. & Giulotto, E. Telomeric repeats far from the ends: Mechanisms of origin and role in evolution. *Cytogenetic and Genome Research* (2009).
35. Martinez, P. *et al.* Mammalian Rap1 controls telomere function and gene expression through binding to telomeric and extratelomeric sites. *Nat. Cell Biol.* (2010).
36. Teo, H. *et al.* Telomere-independent Rap1 is an IKK adaptor and regulates NF-κB-dependent gene expression. *Nat. Cell Biol.* (2010).
37. Simonet, T. *et al.* The human TTAGGG repeat factors 1 and 2 bind to a subset of interstitial telomeric sequences and satellite repeats. *Cell Res.* (2011).
38. Yang, D. *et al.* Human telomeric proteins occupy selective interstitial sites. *Cell Res.* (2011).
39. Wood, A. M. *et al.* TRF2 and lamin A/C interact to facilitate the functional organization of chromosome ends. *Nat. Commun.* (2014).
40. Garrobo, I., Marión, R. M., Domínguez, O., Pisano, D. G. & Blasco, M. A. Genome-wide analysis of in vivo TRF1 binding to chromatin restricts its location exclusively to telomeric repeats. *Cell Cycle* (2014).
41. Watson, J. D. Origin of concatemeric T7 DNA. *Nat. New Biol.* (1972).
42. Olovnikov, A. M. A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *J. Theor. Biol.* (1973).
43. Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. & Sugino, A. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc. Natl. Acad. Sci. U. S. A.* (1968).
44. Harley, C. B., Futcher, A. B. & Greider, C. W. Telomeres shorten during ageing of human fibroblasts. *Nature* (1990).
45. Hastie, N. D. *et al.* Telomere reduction in human colorectal carcinoma and with ageing. *Nature* (1990).
46. Ohki, R., Tsurimoto, T. & Ishikawa, F. In Vitro Reconstitution of the End Replication Problem. *Mol. Cell. Biol.* (2001).
47. Huffman, K. E., Levene, S. D., Tesmer, V. M., Shay, J. W. & Wright, W. E. Telomere shortening is proportional to the size of the G-rich telomeric 3'-overhang. *J. Biol. Chem.* (2000).
48. Longhese, M. P., Bonetti, D., Manfrini, N. & Clerici, M. Mechanisms and regulation of DNA end resection. *EMBO J.* (2010).
49. Wu, P., Takai, H. & De Lange, T. Telomeric 3′ overhangs derive from resection by Exo1 and apollo and fill-in by POT1b-associated CST. *Cell* (2012).
50. Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in tetrahymena extracts. *Cell* (1985).
51. Blackburn, E. H. *et al.* Recognition and elongation of telomeres by telomerase. in *Genome* (1989).
52. Cohen, S. B. *et al.* Protein composition of catalytically active human telomerase from immortal cells. *Science* (2007).
53. Mitchell, J. R., Wood, E. & Collins, K. A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* (1999).
54. Pogacic, V., Dragon, F. & Filipowicz, W. Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10. *Mol. Cell. Biol.* (2000).
55. Nguyen, T. H. D. *et al.* Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* (2018).
56. Autexier, C. & Lue, N. F. The Structure and Function of Telomerase Reverse Transcriptase. *Annu. Rev. Biochem.* (2006).
57. Wyatt, H. D. M., West, S. C. & Beattie, T. L. InTERTpreting telomerase structure and function. *Nucleic Acids Res.* (2010).

58. Flores, I., Cayuela, M. L. & Blasco, M. A. Effects of telomerase and telomere length on epidermal stem cell behavior. *Science* (2005).
59. Deng, Y. & Chang, S. Role of telomeres and telomerase in genomic instability, senescence and cancer. *Laboratory Investigation* (2007).
60. Hiyama, E. & Hiyama, K. Telomere and telomerase in stem cells. *British Journal of Cancer* (2007).
61. Shay, J. W. & Bacchetti, S. A survey of telomerase activity in human cancer. *Eur. J. Cancer* (1997).
62. Shay, J. W. & Wright, W. E. Telomeres and telomerase in normal and cancer stem cells. *FEBS Letters* (2010).
63. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* (2013).
64. Bryan, T. M., Englezou, A., Dalla-Pozza, L., Dunham, M. A. & Reddel, R. R. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat Med* (1997).
65. Cesare, A. J. & Reddel, R. R. Alternative lengthening of telomeres: Models, mechanisms and implications. *Nature Reviews Genetics* (2010).
66. Cho, N. W., Dilley, R. L., Lampson, M. A. & Greenberg, R. A. Interchromosomal homology searches drive directional ALT telomere movement and synapsis. *Cell* (2014).
67. Dunham, M. A., Neumann, A. A., Fasching, C. L. & Reddel, R. R. Telomere maintenance by recombination in human cells. *Nat. Genet.* (2000).
68. Bryan, T. M., Englezou, A., Gupta, J., Bacchetti, S. & Reddel, R. R. Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J* (1995).
69. Muntoni, A. & Reddel, R. R. The first molecular details of ALT in human tumor cells. *Hum. Mol. Genet.* (2005).
70. Apte, M. S. & Cooper, J. P. Life and cancer without telomerase: ALT and other strategies for making sure ends (don't) meet. *Critical Reviews in Biochemistry and Molecular Biology* (2017).
71. Henson, J. D., Neumann, A. A., Yeager, T. R. & Reddel, R. R. Alternative lengthening of telomeres in mammalian cells. *Oncogene* (2002).
72. Liu, L. *et al.* Telomere lengthening early in development. *Nat. Cell Biol.* (2007).
73. Zalzman, M. *et al.* Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature* (2010).
74. Schoorlemmer, J., Pérez-Palacios, R., Climent, M., Guallar, D. & Muniesa, P. Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and Epigenetic Control of Stem Cell Potency. (2014).
75. Makarov, V. L., Lejnine, S., Bedoyan, J. & Langmore, J. P. Nucleosomal organization of telomere-specific chromatin in rat. *Cell* (1993).
76. Tommerup, H., Dousmanis, A. & de Lange, T. Unusual chromatin in human telomeres. *Mol. Cell. Biol.* (1994).
77. Blasco, M. A. The epigenetic regulation of mammalian telomeres. *Nature Reviews Genetics* (2007).
78. Rusche, L. N., Kirchmaier, A. L. & Rine, J. The Establishment, Inheritance, and Function of Silenced Chromatin in Saccharomyces cerevisiae . *Annu. Rev. Biochem.* (2003).
79. Benetti, R., García-Cao, M. & Blasco, M. A. Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. *Nat. Genet.* (2007).
80. García-Cao, M., O'Sullivan, R., Peters, A. H. F. M., Jenuwein, T. & Blasco, M. A. Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nat. Genet.* (2004).
81. Gonzalo, S. *et al.* DNA methyltransferases control telomere length and telomere recombination in mammalian cells. *Nat. Cell Biol.* (2006).
82. Tang, J. *et al.* G-quadruplex preferentially forms at the very 3′ end of vertebrate telomeric DNA. *Nucleic Acids Res.* (2008).
83. Grunstein, M. & Gasser, S. M. Epigenetics in Saccharomyces cerevisiae. *Cold Spring Harb. Perspect. Biol.* (2013).
84. Ichikawa, Y., Nishimura, Y., Kurumizaka, H. & Shimizu, M. Nucleosome organization and chromatin dynamics in telomeres. *Biomolecular Concepts* (2015).
85. Wood, A., Schneider, J. & Shilatifard, A. Cross-talking histones: implications for the regulation of gene expression and DNA repair. *Biochem. Cell Biol.* (2005).
86. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell*

Research (2011).
87. Greer, E. L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* (2012).
88. Di Lorenzo, A. & Bedford, M. T. Histone arginine methylation. *FEBS Letters* (2011).
89. Musselman, C. A., Lalonde, M. E., Côté, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nature Structural and Molecular Biology* (2012).
90. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* (2007).
91. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: Lessons from professional pocket pickers. *Nature Structural and Molecular Biology* (2007).
92. Tardat, M. & Déjardin, J. Telomere chromatin establishment and its maintenance during mammalian development. *Chromosoma* (2018).
93. Cubiles, M. D. *et al.* Epigenetic features of human telomeres. *Nucleic Acids Res.* (2018).
94. Udugama, M. *et al.* Histone variant H3.3 provides the heterochromatic H3 lysine 9 tri-methylation mark at telomeres. *Nucleic Acids Res.* (2015).
95. Farooq, Z., Banday, S., Pandita, T. K. & Altaf, M. The many faces of histone H3K79 methylation. *Mutation Research - Reviews in Mutation Research* (2016).
96. Jamieson, K. *et al.* Telomere repeats induce domains of H3K27 methylation in Neurospora. *Elife* (2018).
97. Jones, B. *et al.* The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. *PLoS Genet.* (2008).
98. Nishibuchi, G. & Déjardin, J. The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals. *Chromosome Research* (2017).
99. Vaquero-Sedas, M. I. & Vega-Palas, M. Á. Targeting Cancer through the Epigenetic Features of Telomeric Regions. *Trends in Cell Biology* (2019).
100. Booth, L. N. & Brunet, A. The Aging Epigenome. *Molecular Cell* (2016).
101. Perrod, S. & Gasser, S. M. Long-range silencing and position effects at telomeres and centromeres: Parallels and differences. *Cellular and Molecular Life Sciences* (2003).
102. Jaiswal, D., Turniansky, R. & Green, E. M. Choose Your Own Adventure: The Role of Histone Modifications in Yeast Cell Fate. *Journal of Molecular Biology* (2017).
103. Benetti, R. *et al.* Suv4-20h deficiency results in telomere elongation and derepression of telomere recombination. *J. Cell Biol.* (2007).
104. Schoeftner, S. & Blasco, M. A. Chromatin regulation and non-coding RNAs at mammalian telomeres. *Semin. Cell Dev. Biol.* (2010).
105. Galati, A., Micheli, E. & Cacchione, S. Chromatin structure in telomere dynamics. *Frontiers in Oncology* (2013).
106. O'Sullivan, R. J., Kubicek, S., Schreiber, S. L. & Karlseder, J. Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. *Nat. Struct. Mol. Biol.* (2010).
107. Rosenfeld, J. A. *et al.* Determination of enriched histone modifications in non-genic portions of the human genome. *BMC Genomics* (2009).
108. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* (2011).
109. Michishita, E. *et al.* SIRT6 is a histone H3 lysine 9 deacetylase that modulates telomeric chromatin. *Nature* (2008).
110. Michishita, E. *et al.* Cell cycle-dependent deacetylation of telomeric histone H3 lysine K56 by human SIRT6. *Cell Cycle* (2009).
111. Canudas, S. *et al.* A role for heterochromatin protein 1γ at human telomeres. *Genes Dev.* (2011).
112. Chow, T. T. *et al.* Local enrichment of HP1alpha at telomeres alters their structure and regulation of telomere protection. *Nat. Commun.* (2018).
113. Mason, J. M., Frydrychova, R. C. & Biessmann, H. Drosophila telomeres: An exception providing new insights. *BioEssays* (2008).
114. Vaquero-Sedas, M. I., Gámez-Arjona, F. M. & Vega-Palas, M. A. Arabidopsis thaliana telomeres exhibit euchromatic features. *Nucleic Acids Res.* (2011).
115. Majerová, E. *et al.* Chromatin features of plant telomeric sequences at terminal vs. internal positions. *Front. Plant Sci.* (2014).
116. Peters, A. H. *et al.* Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* (2001).
117. Schotta, G. *et al.* A chromatin-wide transition to H4K20 monomethylation impairs genome

integrity and programmed DNA rearrangements in the mouse. *Genes Dev.* (2008).
118. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* (2001).
119. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. S. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* (2001).
120. Peters, A. H. F. M. *et al.* Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. *Mol. Cell* (2003).
121. Schotta, G. *et al.* A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* (2004).
122. García-Cao, M., Gonzalo, S., Dean, D. & Blasco, M. A. A role for the Rb family of proteins in controlling telomere length. *Nat. Genet.* (2002).
123. Gonzalo, S. & Blasco, M. A. Role of Rb family in the epigenetic definition of chromatin. *Cell Cycle* (2005).
124. Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* (1995).
125. Lipinski, M. M. & Jacks, T. The retinoblastoma gene family in differentiation and development. *Oncogene* (1999).
126. Classon, M. & Harlow, E. The retinoblastoma tumour suppressor in development and cancer. *Nature Reviews Cancer* (2002).
127. Tasselli, L. *et al.* SIRT6 deacetylates H3K18ac at pericentric chromatin to prevent mitotic errors and cellular senescence. *Nat. Struct. Mol. Biol.* (2016).
128. Brock, G. J. R., Charlton, J. & Bird, A. Densely methylated sequences that are preferentially localized at telomere-proximal regions of human chromosomes. *Gene* (1999).
129. Yehezkel, S., Segev, Y., Viegas-Péquignot, E., Skorecki, K. & Selig, S. Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. *Hum. Mol. Genet.* (2008).
130. Benetti, R. *et al.* A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. *Nat. Struct. Mol. Biol.* (2008).
131. Jiang, W.-Q. *et al.* Induction of alternative lengthening of telomeres-associated PML bodies by p53/p21 requires HP1 proteins. *J. Cell Biol.* (2009).
132. Tennen, R. I., Bua, D. J., Wright, W. E. & Chua, K. F. SIRT6 is required for maintenance of telomere position effect in human cells. *Nat. Commun.* (2011).
133. Van Overveld, P. G. M. *et al.* Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet.* (2003).
134. Pedram, M. *et al.* Telomere Position Effect and Silencing of Transgenes near Telomeres in the Mouse. *Mol. Cell. Biol.* (2006).
135. Steinert, S., Shay, J. W. & Wright, W. E. Modification of Subtelomeric DNA. *Mol. Cell. Biol.* (2004).
136. Benetti, R., Schoeftner, S., Muñoz, P. & Blasco, M. A. Role of TRF2 in the assembly of telomeric chromatin. *Cell Cycle* (2008).
137. Baur, J. A., Zou, Y., Shay, J. W. & Wright, W. E. Telomere position effect in human cells. *Science (80-.).* (2001).
138. Koering, C. E. *et al.* Human telomeric position effect is determined by chromosomal context and telomeric chromatin integrity. *EMBO Rep.* (2002).
139. Stadler, G. *et al.* Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy. *Nat. Struct. Mol. Biol.* (2013).
140. Robin, J. D. *et al.* SORBS2 transcription is activated by telomere position effect-over long distance upon telomere shortening in muscle cells from patients with facioscapulohumeral dystrophy. *Genome Res.* (2015).
141. Sundquist, W. I. & Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* (1989).
142. Williamson, J. R., Raghuraman, M. K. & Cech, T. R. Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell* **59**, 871–880 (1989).
143. Lipps, H. J. & Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends in Cell Biology* (2009).
144. Goytisolo, F. A. *et al.* Short telomeres result in organismal hypersensitivity to ionizing radiation in mammals. *J. Exp. Med.* (2000).
145. Greider, C. W. Telomeres do D-loop-T-loop. *Cell* (1999).
146. Rudenko, G. & Van der Ploeg, L. H. Transcription of telomere repeats in protozoa. *EMBO*

J. (1989).
147. Luke, B. *et al.* The Rat1p 5′ to 3′ Exonuclease Degrades Telomeric Repeat-Containing RNA and Promotes Telomere Elongation in Saccharomyces cerevisiae. *Mol. Cell* (2008).
148. Bah, A., Wischnewski, H., Shchepachev, V. & Azzalin, C. M. The telomeric transcriptome of Schizosaccharomyces pombe. *Nucleic Acids Res.* (2012).
149. Greenwood, J. & Cooper, J. P. Non-coding telomeric and subtelomeric transcripts are differentially regulated by telomeric and heterochromatin assembly factors in fission yeast. *Nucleic Acids Res.* (2012).
150. Vrbsky, J. *et al.* siRNA-mediated methylation of Arabidopsis Telomeres. *PLoS Genet.* (2010).
151. Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* (2007).
152. Schoeftner, S. & Blasco, M. A. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat. Cell Biol.* (2008).
153. Azzalin, C. M. & Lingner, J. Telomere functions grounding on TERRA firma. *Trends in Cell Biology* (2015).
154. Porro, A., Feuerhahn, S., Reichenbach, P. & Lingner, J. Molecular Dissection of Telomeric Repeat-Containing RNA Biogenesis Unveils the Presence of Distinct and Multiple Regulatory Pathways. *Mol. Cell. Biol.* (2010).
155. Deng, Z. *et al.* A role for CTCF and cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end protection. *EMBO J.* (2012).
156. Beishline, K. *et al.* CTCF driven TERRA transcription facilitates completion of telomere DNA replication. *Nat. Commun.* (2017).
157. Nergadze, S. G. *et al.* CpG-island promoters drive transcription of human telomeres. *RNA* (2009).
158. Farnung, B. O., Brun, C. M., Arora, R., Lorenzi, L. E. & Azzalin, C. M. Telomerase efficiently elongates highly transcribing telomeres in human cancer cells. *PLoS One* (2012).
159. Arnoult, N., Van Beneden, A. & Decottignies, A. Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1α. *Nat. Struct. Mol. Biol.* (2012).
160. Porro, A. *et al.* Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat. Commun.* (2014).
161. Lovejoy, C. A. *et al.* Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of Telomeres pathway. *PLoS Genet.* (2012).
162. Episkopou, H. *et al.* Alternative Lengthening of Telomeres is characterized by reduced compaction of telomeric chromatin. *Nucleic Acids Res.* (2014).
163. Arora, R. *et al.* RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat. Commun.* (2014).
164. Chawla, R. & Azzalin, C. M. The telomeric transcriptome and SMG proteins at the crossroads. *Cytogenetic and Genome Research* (2009).
165. De Silanes, I. L., D'Alcontres, M. S. & Blasco, M. A. TERRA transcripts are bound by a complex array of RNA-binding proteins. *Nat. Commun.* (2010).
166. Redon, S., Reichenbach, P. & Lingner, J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res.* (2010).
167. Chu, H. P. *et al.* TERRA RNA Antagonizes ATRX and Protects Telomeres. *Cell* (2017).
168. Cusanelli, E., Romero, C. A. P. & Chartrand, P. Telomeric Noncoding RNA TERRA Is Induced by Telomere Shortening to Nucleate Telomerase Molecules at Short Telomeres. *Mol. Cell* (2013).
169. Yu, T.-Y., Kao, Y. & Lin, J.-J. Telomeric transcripts stimulate telomere recombination to suppress senescence in cells lacking telomerase. *Proc. Natl. Acad. Sci. U. S. A.* (2014).
170. Balk, B., Dees, M., Bender, K. & Luke, B. The differential processing of telomeres in response to increased telomeric transcription and RNA-DNA hybrid accumulation. *RNA Biology* (2014).
171. Flynn, R. L. *et al.* TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature* (2011).
172. De Silanes, I. L. *et al.* Identification of TERRA locus unveils a telomere protection role through association to nearly all chromosomes. *Nat. Commun.* (2014).

173. Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. & Lieberman, P. M. TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres. *Mol. Cell* (2009).
174. Lee, Y. W., Arora, R., Wischnewski, H. & Azzalin, C. M. TRF1 participates in chromosome end protection by averting TRF2-dependent telomeric R loops. *Nat. Struct. Mol. Biol.* (2018).
175. García-Rubio, M. *et al.* Yra1-bound RNA–DNA hybrids cause orientation-independent transcription– replication collisions and telomere instability. *Genes Dev.* (2018).
176. Hirashima, K. & Seimiya, H. Telomeric repeat-containing RNA/G-quadruplex-forming sequences cause genome-wide alteration of gene expression in human cancer cells in vivo. *Nucleic Acids Res.* (2015).
177. Pfeiffer, V. & Lingner, J. TERRA promotes telomere shortening through exonuclease 1-mediated resection of chromosome ends. *PLoS Genet.* (2012).
178. Chu, H. P. *et al.* PAR-TERRA directs homologous sex chromosome pairing. *Nat. Struct. Mol. Biol.* (2017).
179. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* (1981).
180. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U. S. A.* (1981).
181. Thomson, J. A. Embryonic stem cell lines derived from human blastocysts. *Science* (1998).
182. Nichols, J. & Smith, A. Naive and Primed Pluripotent States. *Cell Stem Cell* (2009).
183. Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J. H. Dynamic stem cell states: Naive to primed pluripotency in rodents and humans. *Nature Reviews Molecular Cell Biology* (2016).
184. Gurdon, J. B. & Melton, D. A. Nuclear reprogramming in cells. *Science* (2008).
185. Hochedlinger, K. & Plath, K. Epigenetic reprogramming and induced pluripotency. *Development* (2009).
186. Rideout, W. M., Eggan, K. & Jaenisch, R. Nuclear cloning and epigenetic reprogramming of the genome. *Science* (2001).
187. Yang, X. & Smith, S. L. ES cells derived from cloned embryos in monkey - A jump toward human therapeutic cloning. *Cell Res.* (2007).
188. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. S. Viable offspring derived from fetal and adult mammalian cells. *Nature* (1997).
189. Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. & Tada, T. Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Curr. Biol.* (2001).
190. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* (2006).
191. Amabile, G. & Meissner, A. Induced pluripotent stem cells: current progress and potential for regenerative medicine. *Trends in Molecular Medicine* (2009).
192. Huang, Y., Liang, P., Liu, D., Huang, J. & Songyang, Z. Telomere regulation in pluripotent stem cells. *Protein and Cell* (2014).
193. Yamanaka, S., Takahashi, K., Okita, K. & Nakagawa, M. Induction of pluripotent stem cells from fibroblast cultures. *Nat. Protoc.* (2007).
194. Drissi, R., Zindy, F., Roussel, M. F. & Cleveland, J. L. c-Myc-mediated Regulation of Telomerase Activity Is Disabled in Immortalized Cells. *J. Biol. Chem.* (2001).
195. Flores, I., Evan, G. & Blasco, M. A. Genetic Analysis of Myc and Telomerase Interactions In Vivo. *Mol. Cell. Biol.* (2006).
196. Wu, K. J. *et al.* Direct activation of TERT transcription by c-MYC. *Nat. Genet.* (1999).
197. Suhr, S. T. *et al.* Telomere dynamics in human cells reprogrammed to pluripotency. *PLoS One* (2009).
198. Wong, C. W. *et al.* Krüppel-like transcription factor 4 contributes to maintenance of telomerase activity in stem cells. *Stem Cells* (2010).
199. Hoffmeyer, K. *et al.* Wnt/ -Catenin Signaling Regulates Telomerase in Stem Cells and Cancer Cells. *Science* (2012).
200. Agarwal, S. *et al.* Telomere elongation in induced pluripotent stem cells from dyskeratosis congenita patients. *Nature* (2010). 2
201. Mathew, R. *et al.* Robust activation of the human but not mouse telomerase gene during the induction of pluripotency. *FASEB J.* (2010).
202. Huang, J. *et al.* Association of telomere length with authentic pluripotency of ES/iPS cells.

Cell Res. (2011).
203. Pucci, F., Gardano, L. & Harrington, L. Short telomeres in ESCs lead to unstable differentiation. *Cell Stem Cell* (2013).
204. Armstrong, L. *et al.* Overexpression of Telomerase Confers Growth Advantage, Stress Resistance, and Enhanced Differentiation of ESCs Toward the Hematopoietic Lineage. *Stem Cells* (2005).
205. Yang, C. *et al.* A Key Role for Telomerase Reverse Transcriptase Unit in Modulating Human Embryonic Stem Cell Proliferation, Cell Cycle Dynamics, and In Vitro Differentiation. *Stem Cells* (2008).
206. Marión, R. M. *et al.* Common Telomere Changes during In Vivo Reprogramming and Early Stages of Tumorigenesis. *Stem Cell Reports* (2017).
207. Utikal, J. *et al.* Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature* (2009).
208. Ginis, I. *et al.* Differences between human and mouse embryonic stem cells. *Dev. Biol.* (2004).
209. Schneider, R. P. *et al.* TRF1 is a stem cell marker and is essential for the generation of induced pluripotent stem cells. *Nat. Commun.*(2013).
210. Karlseder, J. *et al.* Targeted deletion reveals an essential function for the telomere length regulator Trf1. *Mol. Cell. Biol.* (2003).
211. Boué, S., Paramonov, I., Barrero, M. J. & Belmonte, J. C. I. Analysis of human and mouse reprogramming of somatic cells to induced pluripotent stem cells. what is in the plate? *PLoS One* (2010).
212. Hosseinpour, B., Bakhtiarizadeh, M. R., Khosravi, P. & Ebrahimie, E. Predicting distinct organization of transcription factor binding sites on the promoter regions: A new genome-based approach to expand human embryonic stem cell regulatory network. *Gene* (2013).
213. Varela, E., Schneider, R. P., Ortega, S. & Blasco, M. A. Different telomere-length dynamics at the inner cell mass versus established embryonic stem (ES) cells. *Proc. Natl. Acad. Sci.* (2011).
214. Arora, R., Brun, C. M. & Azzalin, C. M. Transcription regulates telomere dynamics in human cancer cells. *RNA* (2012).
215. Kar, A., Willcox, S. & Griffith, J. D. Transcription of telomeric DNA leads to high levels of homologous recombination and t-loops. *Nucleic Acids Res.* (2016).
216. Avogaro, L. *et al.* Live-cell imaging reveals the dynamics and function of single-telomere TERRA molecules in cancer cells. *RNA Biol.* (2018).
217. Stong, N. *et al.* Subtelomeric CTCF and cohesin binding site organization using improved subtelomere assemblies and a novel annotation pipeline. *Genome Res.* (2014).
218. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* (2013).
219. Cong, L. & Zhang, F. Genome engineering using crispr-cas9 system. in *Chromosomal Mutagenesis: Second Edition* (2014).
220. Phan, A. T. Human telomeric G-quadruplex: Structures of DNA and RNA sequences. *FEBS Journal* (2010).
221. Koskas, S. *et al.* Heat shock factor 1 promotes TERRA transcription and telomere protection upon heat stress. *Nucleic Acids Res.* (2017).
222. Tutton, S. *et al.* Subtelomeric p53 binding prevents accumulation of DNA damage at human telomeres . *EMBO J.* (2016).
223. Mikulski, P., Komarynets, O., Fachinelli, F., Weber, A. P. M. & Schubert, D. Characterization of the polycomb-group mark H3K27me3 in unicellular algae. *Front. Plant Sci.* (2017).
224. Klocko, A. D. *et al.* Normal chromosome conformation depends on subtelomeric facultative heterochromatin in Neurospora crassa. *Proc. Natl. Acad. Sci. U. S. A.* (2016).
225. Pengelly, A. R., Copur, Ö., Jäckle, H., Herzig, A. & Müller, J. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor polycomb. *Science* (2013).
226. Kotake, Y. *et al.* Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15 INK4B tumor suppressor gene. *Oncogene* (2011).
227. Gutschner, T., Hämmerle, M. & Diederichs, S. MALAT1 - A paradigm for long noncoding RNA function in cancer. *Journal of Molecular Medicine* (2013).
228. Rinn, J. L. *et al.* Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* (2007).
229. Brockdorff, N. Polycomb complexes in X chromosome inactivation. *Philosophical*

Transactions of the Royal Society B: Biological Sciences (2017).
230. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science (80-.).* (2008).
231. Wang, X. *et al.* Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. *Mol. Cell* (2017).
232. Sarma, K. *et al.* ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. *Cell* (2014).
233. Flynn, R. L. *et al.* Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. *Science (80-.).* (2015). doi:10.1126/science.1257216
234. Lin, T. & Lin, Y. p53 switches off pluripotency on differentiation. *Stem cell research & therapy* (2017).
235. Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* (2009).
236. Li, H. *et al.* The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* (2009).
237. Marión, R. M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* (2009).
238. Xu, K. *et al.* EZH2 Oncogenic Activity in Castration-Resistant Prostate Cancer Cells Is Polycomb-Independent. *Science (80-.).* (2012).
239. Shi, B. *et al.* Integration of Estrogen and Wnt Signaling Circuits by the Polycomb Group Protein EZH2 in Breast Cancer Cells. *Mol. Cell. Biol.* (2007).
240. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* (2006).
241. Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* (2006).
242. Pan, G. *et al.* Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. *Cell Stem Cell* (2007).
243. Zhao, X. D. *et al.* Whole-Genome Mapping of Histone H3 Lys4 and 27 Trimethylations Reveals Distinct Genomic Compartments in Human Embryonic Stem Cells. *Cell Stem Cell* (2007).
244. Maherali, N. *et al.* Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution. *Cell Stem Cell* (2007).
245. Guenther, M. G. *et al.* Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell* (2010).
246. Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* (2012).
247. Zhou, Y., Hartwig, B., James, G. V., Schneeberger, K. & Turck, F. Complementary activities of TELOMERE REPEAT BINDING proteins and polycomb group complexes in transcriptional regulation of target genes. *Plant Cell* (2016).
248. Zhou, Y. *et al.* Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in Arabidopsis. *Nat. Genet.* (2018).
249. Zeng, S., Liu, L., Sun, Y., Lu, G. & Lin, G. Role of telomeric repeat-containing RNA in telomeric chromatin remodeling during the early expansion of human embryonic stem cells. *FASEB J.* (2017).
250. Sadhukhan, R., Chowdhury, P., Ghosh, S. & Ghosh, U. Expression of Telomere-Associated Proteins is Interdependent to Stabilize Native Telomere Structure and Telomere Dysfunction by G-Quadruplex Ligand Causes TERRA Upregulation. *Cell Biochem. Biophys.* (2018).
251. Bejarano, L. *et al.* Inhibition of TRF1 Telomere Protein Impairs Tumor Initiation and Progression in Glioblastoma Mouse Models and Patient-Derived Xenografts. *Cancer Cell* (2017).
252. García-Beccaria, M. *et al.* Therapeutic inhibition of TRF1 impairs the growth of p53-deficient K-RasG12V-induced lung cancer by induction of telomeric DNA damage. *EMBO Mol. Med.* (2015).
253. Wang, L. *et al.* Dual roles of TRF1 in tethering telomeres to the nuclear envelope and protecting them from fusion during meiosis. *Cell Death Differ.* (2018).

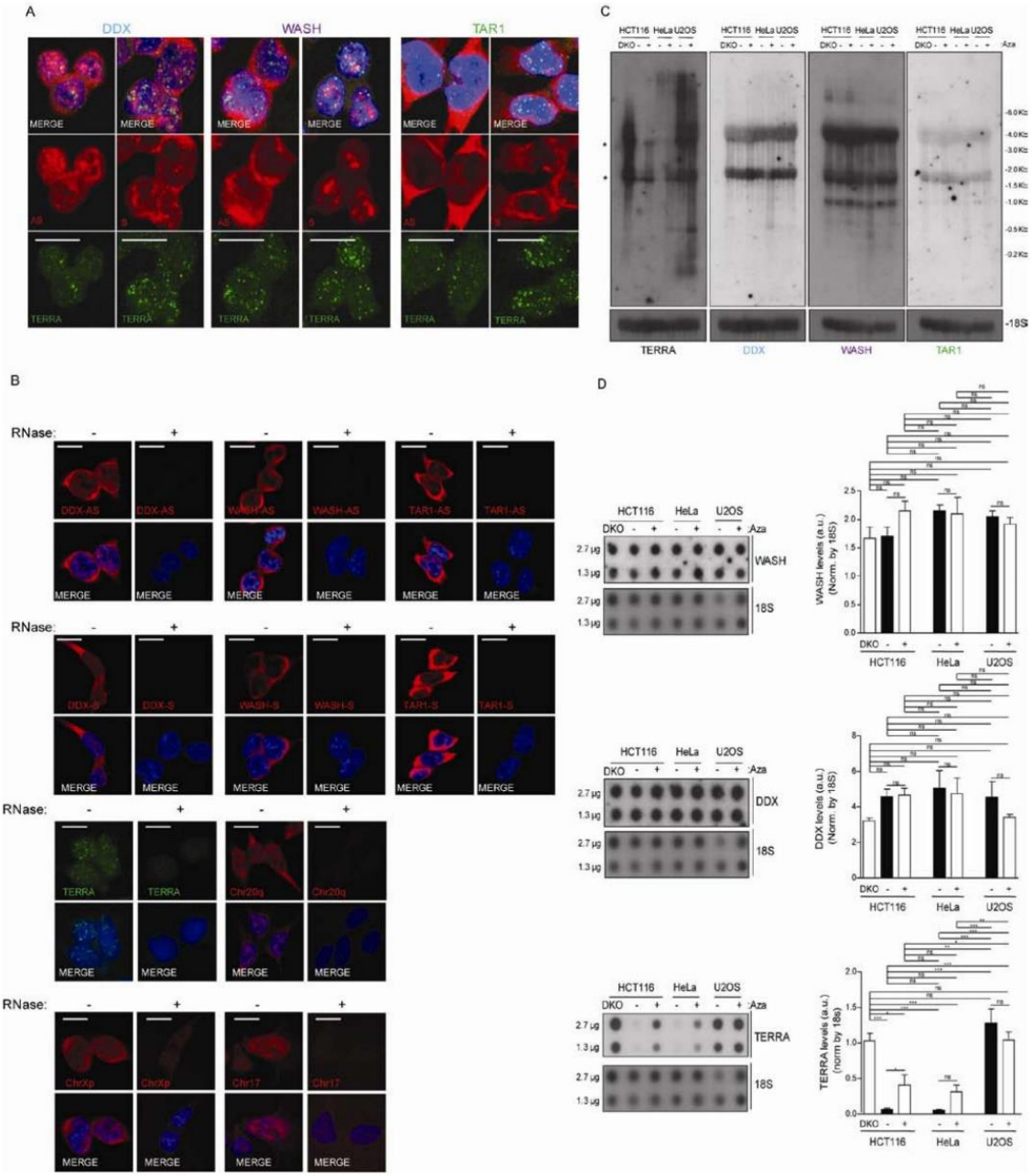
Anexos

Anexo 1

Figuras suplementarias de los artículos principales que componen esta tesis doctoral

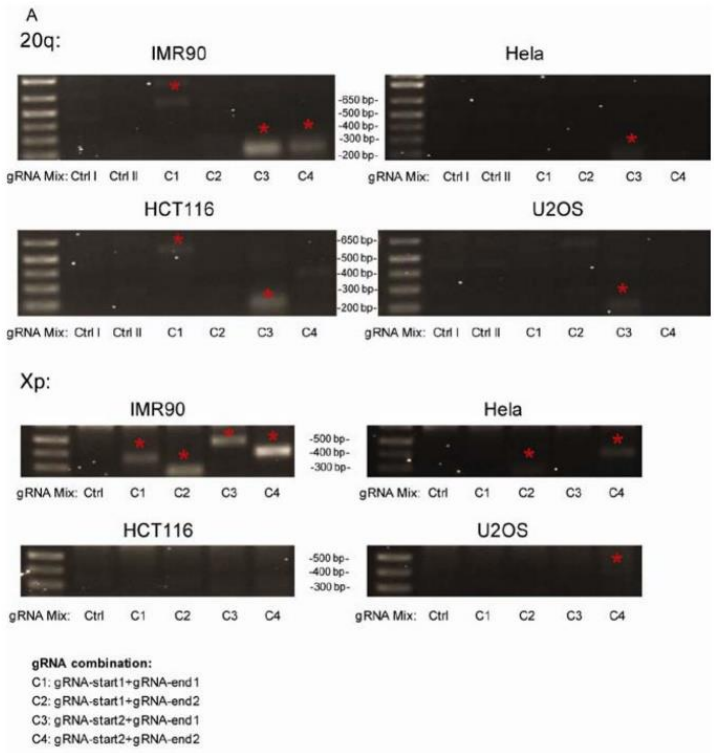
Anexo 1.1: Telomeric RNAs are essential to maintain telomeres

Supplementary Figures

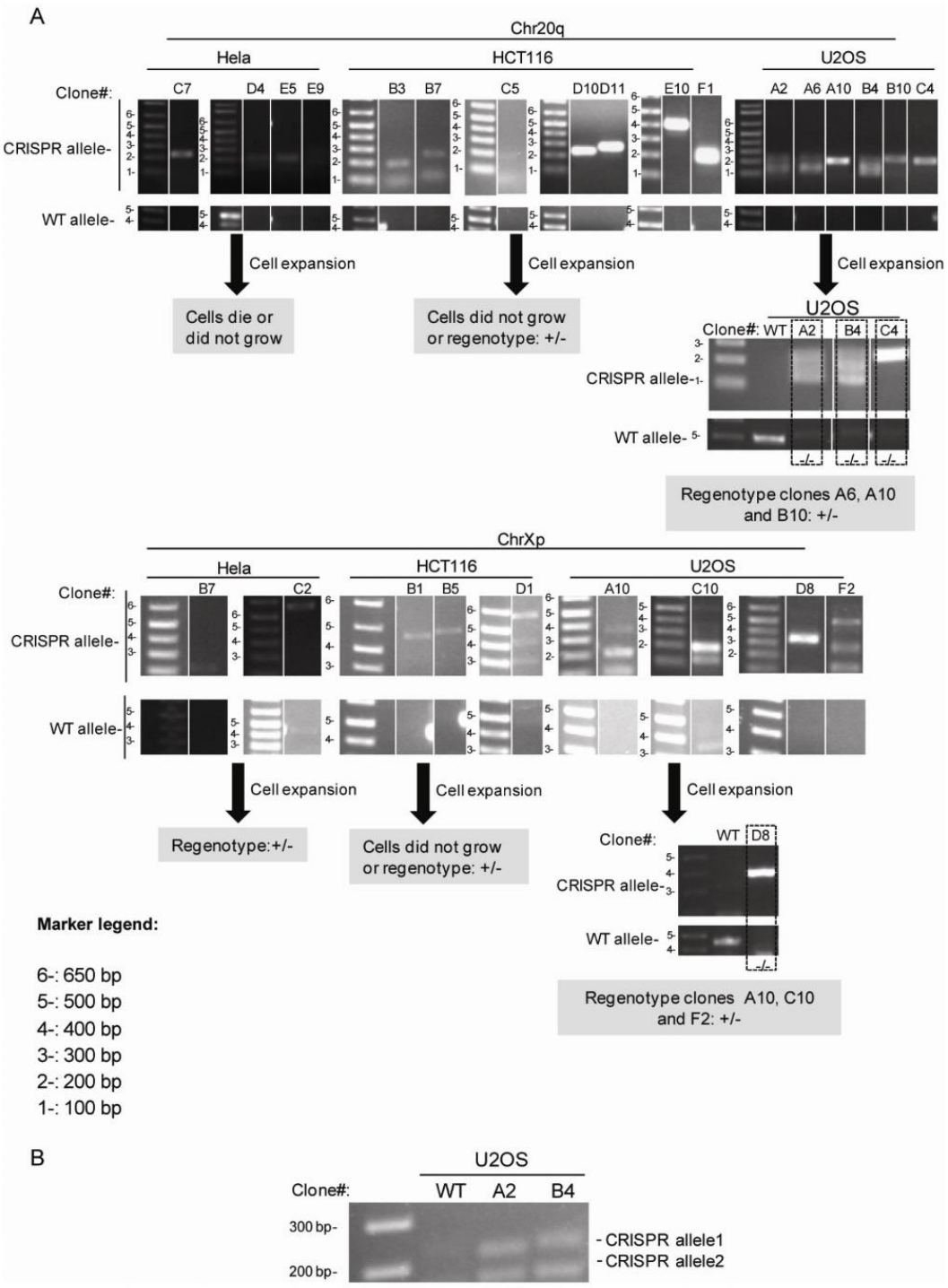


Montero et al._Supplementary Figure 1

Supplementary Figure 1. Transcripts arising from the structurally conserved subtelomeres do not show TERRA features. (A) Confocal microscopy images of double RNA-FISH using probes targeting either TAR1, DDX11L or WASH transcripts (red) and TERRA-UUAGGG track (green). At least 50 nuclei were analyzed in all the cases. AS: probe complementary to the transcript arising from the positive DNA strand; S: probe complementary to the transcript from the negative DNA strand. Scale bar: 10 μm. **(B)** Confocal microscopy images from RNA-FISH with the different RNA-FISH probes in the absence or presence of RNase. Scale bar: 10 μm. **(C)** Northern blotting using ^{32}P-dCTP-labelled probes on total RNA from HCT116 cells DKO (for the DNMT1 and 3b genes) or HCT116 wild-type, HeLa and U2OS treated and untreated with 5'Azacytidin (Aza) for 72 hours. The RNA was used to prepare two identical membranes, in one of them the TERRA probe was hybridized first and, upon stripping, the WASH probe. In the other membrane, the DDX probe was hybridized first and then the TAR1 probe. 18S was included as a loading control. The expected size for DDX transcripts is 1.4-1.6kb and for WASH 1.3-1.8Kb. For TAR1 region, no annotated transcripts have been described so far. *Unspecific band due to cross-hybridization with rRNA 18S and 28S. **(D)** The RNA extracted in (B) was also used for RNA dot-blot to detect either TERRA-UUAGGG track, DDX11L or WASH transcripts; 18S serves as loading control. All the hybridizations were performed on the same membrane one after the other upon stripping. (Graphs) Dot-blot quantification normalized by 18S (mean values±s.e.m., n=3 biological replicates). The Student´s t-test was used for the statistical analysis of the comparison between untreated and Aza-treated cells. One-way Anova with Tukey post test was used for the statistical analysis of the comparison of one cell line with the others (*$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$).

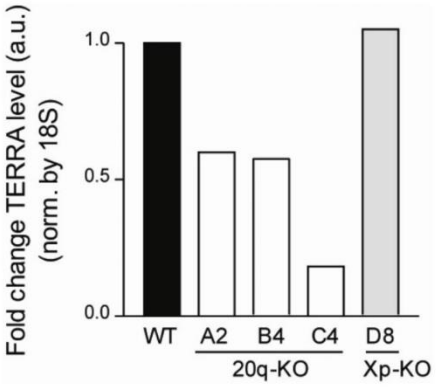


Supplementary Figure 2. Different combinations of gRNAs are able to delete the 20q and Xp loci in different cell lines. IMR90 non-transformed human fibroblasts, HeLa, HCT166 and U2OS cell lines were transfected with different combinations of Cas9-GFP-plasmids containing two different gRNAs to delete 10Kb of the 20q and XpYp region enriched in TERRA transcripts. Two-days upon transfection total DNA was isolated from the different cells lines and PCR using specifics primers to detect the deletion was performed and visualized in ethidium bromide gels. The individual gRNAs used in each combination (C1, C2, C3 and C4) (gRNA mix) are shown below. The primers for the gRNAs synthesis can be found in the **Supplementary Table 5**. In both cases control PCRs (Ctrl I and Ctrl II) using DNA for non treated cells are shown. * DNA amplification found in the treated cells PCR but not in the control PCR.

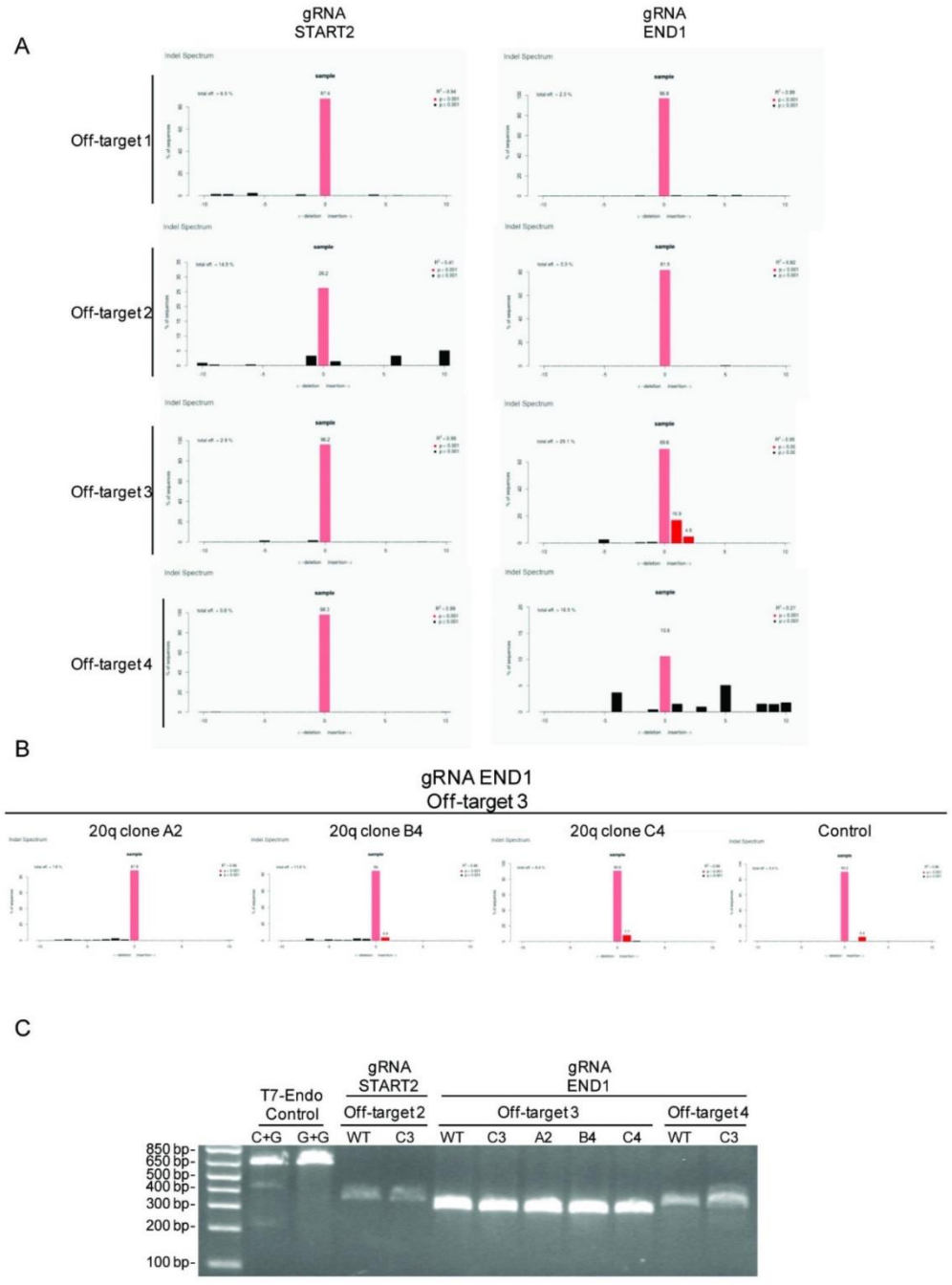


Montero et al._Supplementary Figure 3

Supplementary Figure 3. Regenotyping of false homozygous clones upon expansion. **(A)** Ethidium bromide gels showing the genotype for the CRISPR and wild-type allele in the different clones of the Hela, HCT116 and U2OS cells obtained upon the deletion of the 20q and Xp loci before and after expansion. The primers to detect the deletion were designed to obtain a PCR product of 219 bp for the 20q deletion and of 407bp for the Xp deletion but other sizes can be expected depending on how the CRISPR cut&repair was performed by the cell. The DNA marker legend is shown at the left bottom. Ethidium bromide gels with different expositions are due to the run of the different samples on different days. The white strips in the gels indicate the removal of irrelevant samples from that gel **(B)** Ethidium bromide gels showing the two different CRISPR alleles in the 20q-KO clones A2 and B4

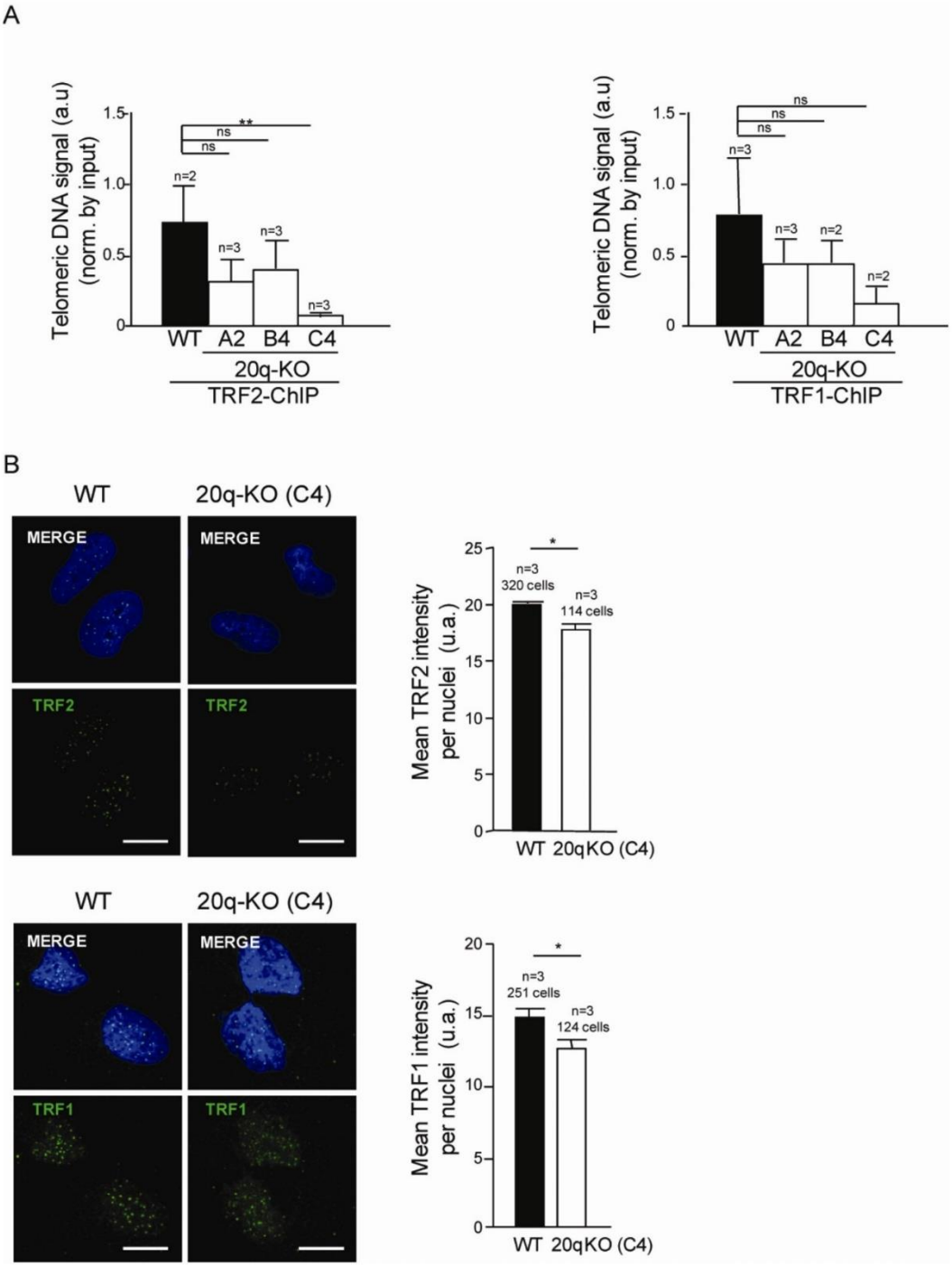


Supplementary Figure 4. Total TERRA levels upon deletion of the 20q locus. Quantification of TERRA signal obtained in the Northern blot from Fig. 3B normalized by 18S. Fold change is shown



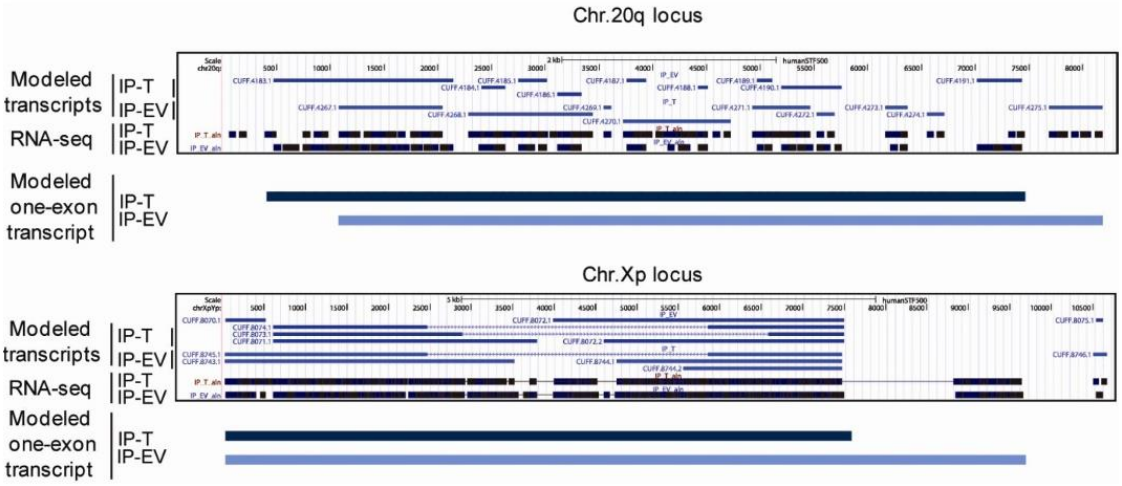
Montero et al._Supplementary Figure 5

Supplementary Figure 5. Off-target mutation analysis in the 20q-TERRA KO cells. (A) TIDE (Tracking of Indels by DEcomposition) analysis to study the four top of targets for the gRNAs (START2 and END1) used to generate the 20q locus deletion. The analysis was performed with DNA pools from either wild type cells or cells bearing the 20q deletion, named C3 (from which the clones A2, B4 and C4 were obtained from). A DNA that did not undergo CRISPR treatment was used as reference. **(B)** TIDE analysis for the off-targets 3, gRNA END1 using DNAs from the 20q-KO clones A2, B4 and C4 and from cells that did not undergo CRSIPR treatment. A DNA from other set of cells that did not undergo CRISPR treatment was used as reference **(C)** T7 endonuclease assay was performed on DNAs from the regions of interest (off-target2/START2, off-target 3/END1 and off-target 4/END1) obtained from pool of cells bearing the 20q deletion (C3) or from the 20q-KO clones. Two identical DNAs except for a C to G mutation in one of them (named C and G) were used as positive control. The ethidium bromide gels shows the different DNAs upon T7 endonuclease digestion.

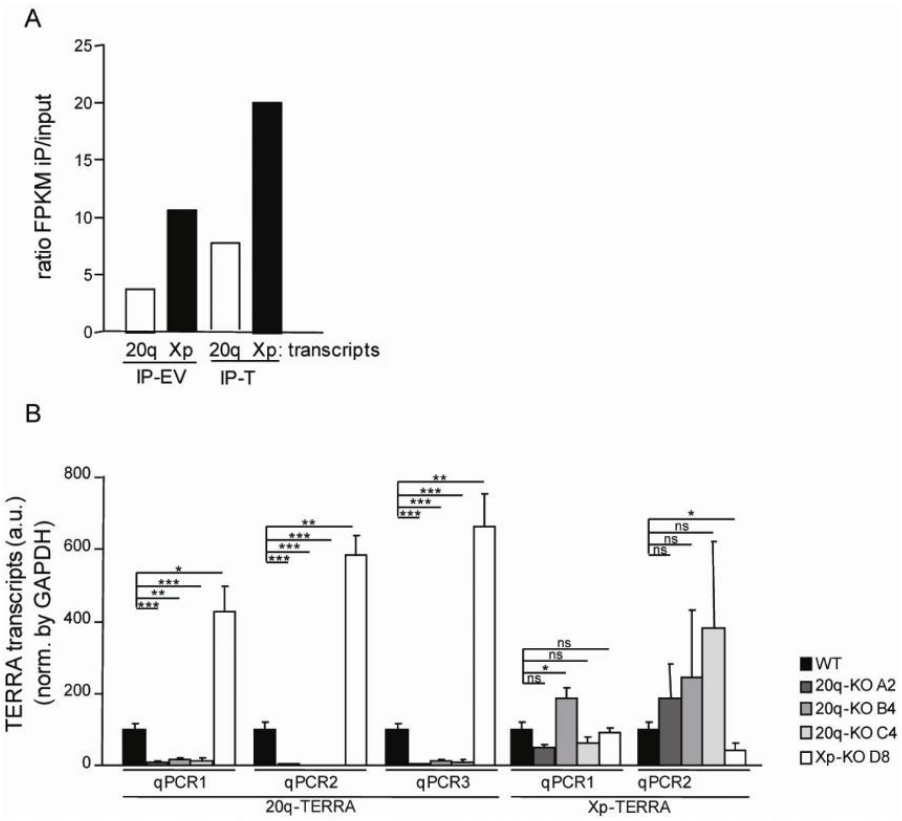


Montero et al._Supplementary Figure 6

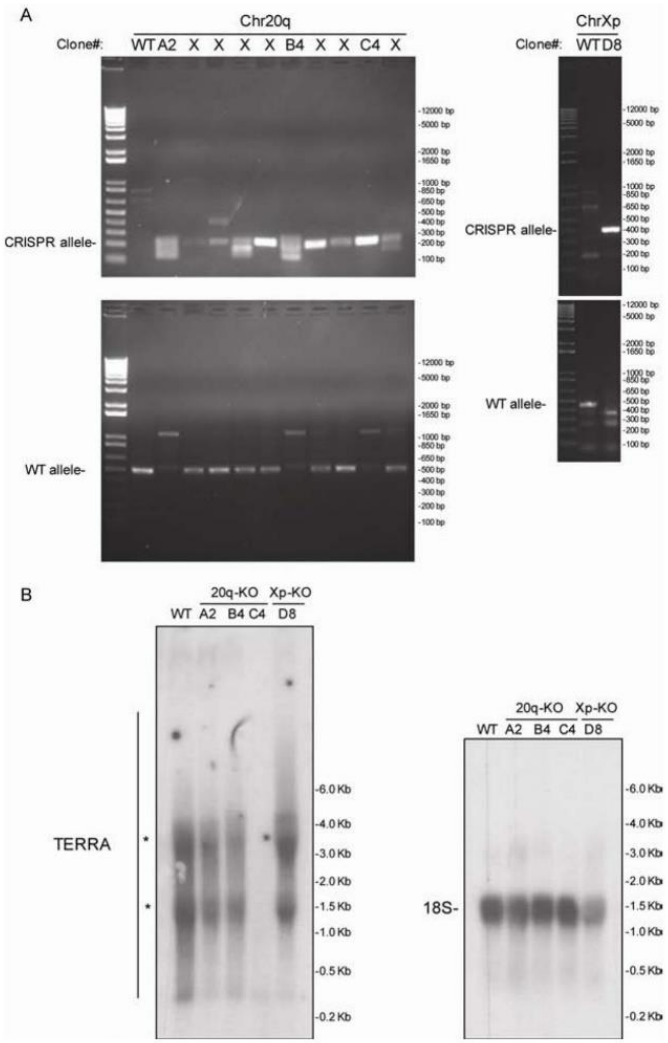
Supplementary Figure 6 Detection of shelterins in 20q-TERRA KO cells. (A) Quantification of immunoprecipitated telomeric repeats upon TRF2- or TRF1-ChIP. Values were obtained after normalization to telomeric DNA input and corrected by the differences in the telomere length between the clones and wild-type cells (see Figure 4). Error bars correspond to the 2-3 independent experiments (mean values± s.e.m., n=2-3 independent experiments). One-way Anova with Tukey post test was used for statistical analysis (**p < 0.01). **(B)** Representative images of TRF2 and TRF1 immunoflorescence (green) in wild-type cells and 20q-KO TERRA cells (clone C4). Merge images with DAPI are also shown (Graphs). Quantification of the mean fluorescence intensity of TRF1 and TRF2 is represented (mean values± s.e.m., n=3 independent experiments). The Student´s t-test was used for statistical analysis (*p < 0.05). Scale bar: 10 μm.



Supplementary Figure 7. Comparison of 20q-TERRA transcripts with Xp-RNAs. Raw data from RNA-seq samples (IP-EV and IP-T)[1] was downloaded and aligned to the human subtelomeric reference genome. Next, transcripts were modeled by Cufflinks in the chr.20q and chr.Xp subtelomeric regions of interest (corresponding to the regions to be deleted with CRISPR/Cas9) and pairwise alignments performed (see Materials and Methods). The snapshot shows (from top to bottom) modeled transcripts in the different IPs, RNA-seq reads (dense format) and the modeled one-exon transcript for each condition



Supplementary Figure 8. Expression analysis of the 20q-TERRA transcripts with respect the Xp-RNAs. (A) Raw data from RNA-seq samples (IP-EV and IP-T) (Porro et al., 2014) was downloaded and aligned to the human subtelomeric reference genome. FPKM (fragments per kilobase of transcript per million fragments mapped) in the chr.20q and chr.Xp subtelomeric regions of interest (corresponding to the regions to be deleted with CRISPR/Cas9) were calculated for each of the IPs and input. (Graph). Quantification of the FPKM in each IP normalized by their corresponding input. **(B)** Detection of the 20q-TERRA or Xp transcripts by PCR using different pair of primers in wild-type cells, 20q-KO clones (A2, B4, C4) and Xp-KO clone (D8). The fold RNA enrichment normalized by GAPDH is shown. Primers 20q/qPCR1-3 are designed within the 20q TERRA locus. Primer Xp/qPCR 1 and 2 are designed within the Xp locus but only the Xp/qPCR2 is designed within the Xp deletion performed in clone D8.



Supplementary Figure 9. Raw gels and blots from Figure 2E and 3B. (A) Ethidium bromide gels showing the CRISPR allele for the deletion of the 20q and Xp loci detected by PCR in different clones of the U2OS cells. The white strips in the gels indicate the removal of irrelevant samples from that gel. 'X' are irrelevant samples. **(B)** Northern blotting using 32P-dCTP-labelled probe against TERRA-UUAGGG tract in the U2OS cells WT or KO for the 20q or Xp loci. 18S was included as a loading control. *Unspecific band due to cross-hybridization with rRNA 18S and 28S.

Supplementary Tables

Supplementary Table 1. Chromosome ends amplified by probes against TAR1, DDX and WASH

Probe	Amplified region
TAR1	1q,2q,4q,10q,21q,22q
DDX	1p,2,9p,12p,15q,16p,Xq,Yq
WASH	1p,9p,12p,Xq,Yq

Supplementary Table 2. TAR1 probe specificity recognizing different chromosome ends

	Max score	Total score	Query cove	E value	Ident	Probe origin
Chr1q	697	697	1%	0	100%	1q
Chr2q	697	697	1%	0	100%	2q,4q
Chr5q	675	675	1%	0	99%	2q,4q
Chr6q	675	675	1%	0	99%	10q
Chr9p	481	481	1%	1.00E-138	90%	10q
Chr10q	697	697	1%	0	100%	10q
Chr13q	664	664	1%	0	98%	10q
Chr15q	464	464	1%	1.00E-133	89%	10q
Chr16p	523	523	1%	2.00E-151	92%	10q
Chr22q	697	697	1%	0	100%	22q
ChrXq	483	483	1%	4.00E-139	90%	10q

Supplementary Table 3. DDX probe specificity recognizing different chromosome ends

	Max score	Total score	Query cove	E value	Ident	Probe origin
Chr1p	710	710	1%	0	100%	1p
Chr9p	710	710	1%	0	100%	9p,Xq,Yq
Chr12p	710	710	1%	0	100%	12p
Chr15q	710	710	1%	0	100%	15q
Chr16p	710	710	1%	0	100%	16p
Chr19p	706	706	1%	0	100%	1p
ChrXq	710	710	1%	0	100%	9p,Xq,Yq

Supplementary Table 4. WASH probe specificity recognizing different chromosome ends

	Max score	Total score	Query cove	E value	Ident	Probe origin
Chr1p	701	701	1%	0	100%	1p
Chr9p	697	697	1%	0	100%	9p

Chr12p	701	701	1%	0	100%	12p
Chr15q	682	682	1%	0	99%	1p
Chr16p	645	645	1%	0	98%	1p
Chr19p	676	676	1%	0	99%	1p
ChrXq	701	701	1%	0	100%	Xq,Yq

Supplementary Table 5. Primers for gRNA cloning

Region	Primer	Sequence (5´-3´)
Chr20q	gRNA20q-start1	CACCgTATAGCGGCGGCACGCCGCC
	R-gRNA20q-start1	AAACGGCGGCGTGCCGCCGCTATAc
	gRNA20q-start2	CACCgCAGGCTGGCGCGACGTGCGG
	R-gRNA20q-start2	AAACCCGCACGTCGCGCCAGCCTGc
	gRNA20q-end1	CACCgCCCACCACCTTAGCGGATCA
	R-gRNA20q-end1	AAACTGATCCGCTAAGGTGGTGGGc
	gRNA20q-end2	CACCgAAATTTTACTTCATCGGAGG
	R-gRNA20q-end2	AAACCCTCCGATGAAGTAAAATTTc
ChrXpYp	gRNAXpYp-start1	CACCgCATGGTGGGGACCCGATGCT
	R-gRNAXpYp-start1	AAACAGCATCGGGTCCCCACCATGc
	gRNAXpYp-start2	CACCGCCACCATATCATACTCGGA
	R-gRNAXpYp-start2	AAACTCCGAGTATGATATGGTGGC
	gRNAXpYp-end1	CACCgAACCGGGTGGGTACGTCAAC
	R-gRNAXpYp-end1	AAACGTTGACGTACCCACCCGGTTc
	gRNAXpYp-end2	CACCGGTCTACACCGGCTTCATCC
	R-gRNAXpYp-end2	AAACGGATGAAGCCGGTGTAGACC

Supplementary Table 6. gRNA combinations

Combination#1: gRNA-start1+gRNA-end1
Combination #2: gRNA-start1+gRNA-end2
Combination #3: gRNA-start2+gRNA-end1
Combination #4: gRNA-start2+gRNA-end2

Supplementary Table 7. Primers used to amplify the KO alelle

gRNA combination	Primer	Sequence (5´-3´)
#1,#2,#3,#4	F-20qKOI	CCCGGGTCTGTGTTAAGC
#1,#3	R-20qKOI	GATCAGGGCTGTCTTAATGC
#2,#4	R-0qKOII	TCATCACACCTGCATATACTGT
#1,#2,#3,#4	F-pYpKO	GGCAAAGGGAGCAGTCAT
#1,#2,#3,#4	R-pYpKO	GACGCAACTTCACAGTTACA

Supplementary Table 8. Primers to detect transcripts from the 20q and Xp loci

Region	Primer	Sequence (5´-3´)
20q	F120q-qpcr	CTGGTGCCAGAGTGGATT
	R120q-qpcr	CACCTGTTCTCTTTGTCTGG
	F220q-qpcr	ACATGGGCGATACTCAGG
	R220q-qpcr	CCCACTACTGTGCCTCAA
	F320q-qpcr	GAAGTTGCTGGGTTCTATGG
	R320q-qpcr	ATGGTGCAGACACTGTGG
Xp	F1xp-qpcr	GCAAAGAGTGAAAGAACGAAGCTT
	R1xp-qpcr	CCCTCTGAAAGTGGACCTATCA
	F2xp-qpcr	CTTGAGCTCTCACCACTCAC
	R2xp-qpcr	GCCACGATAGCTTCTTCC

Supplementary Table 9. Chromosomal rearrangements found by CGH array

Chr	Cytoband	Start	Stop	Gain	Loss	Deletion	pval
chr3	p26.3 - p26.1	212.711	6.063.107	0,000000	-0,495877	0,000000	6.11E-97
chr4	q24 - q28.1	102.668.388	126.615.872	0,000000	-0,273520	0,000000	7.73E-09
chr6	p25.3 - p12.3	407.231	51.356.459	0,000000	-0,540169	0,000000	4,900e-324
chr6	p25.3 - p22.2	439.115	26.104.535	0,000000	-0,775369	0,000000	1.51E-80
chr6	p25.3 - p24.3	685.345	9.826.230	0,000000	-0,907387	0,000000	5.65E-07
chr6	p22.1 - p12.3	27.049.061	49.895.377	0,000000	-0,417303	0,000000	6.61E-42
chr6	q27	166.907.999	170.890.108	0,365987	0,000000	0,000000	1.07E-43
chr7	p22.3 - p21.3	707.018	12.258.124	0,313862	0,000000	0,000000	3.10E-92
chr7	p21.3	8.153.319	12.193.158	0,506460	0,000000	0,000000	5.47E-08
chr9	p24.3	329.684	1.411.809	0,000000	-0,866823	0,000000	1.43E-42
chr9	p24.3 - p21.3	1.869.792	23.690.332	0,778509	0,000000	0,000000	4,900e-324
chr9	p24.3 - p24.2	1.869.792	4.357.696	0,517094	0,000000	0,000000	9.67E-11
chr9	p24.2 - p23	4.516.000	11.573.590	1,181,035	0,000000	0,000000	1.82E-62
chr9	p22.1 - p21.3	18.838.925	23.660.030	0,502066	0,000000	0,000000	5.72E-29
chr10	p15.2 - p15.1	3.735.020	4.133.505	0,000000	0,000000	-1,917,017	1.37E-27
chr10	q11.21 - q24.1	44.877.427	97.315.278	0,000000	-0,359404	0,000000	4,900e-324
chr10	q23.1 - q23.33	84.298.588	96.872.423	0,000000	-0,474244	0,000000	2.15E-11
chr12	p11.22 - p11.1	29.329.809	34.345.585	0,000000	-0,401302	0,000000	1.07E-61
chr12	p11.1	33.636.657	34.345.585	0,000000	0,000000	-1,189,862	1.05E-23
chr14	q11.2 - q24.3	20.253.739	78.177.244	0,308504	0,000000	0,000000	4,900e-324
chr14	q11.2 - q21.2	21.697.792	45.083.729	0,411750	0,000000	0,000000	1.20E-17
chr15	q12 - q13.1	27.571.940	28.525.460	0,000000	-0,334935	0,000000	1.24E-09
chr15	q14 - q26.3	38.466.370	98.966.441	0,000000	-0,461232	0,000000	4,900e-324
chr15	q14 - q22.2	39.979.988	60.823.012	0,000000	-0,284430	0,000000	4.02E-65
chr15	q23 - q26.2	71.272.476	96.982.260	0,000000	-0,642954	0,000000	7.53E-74
chr15	q23 - q24.2	72.670.081	75.224.254	0,000000	-0,877269	0,000000	3.38E-12
chr18	p11.32 -p11.21	142.096	13.971.521	0,000000	-0,571498	0,000000	4,900e-324

chr18	p11.32	142.096	293.595	0,000000	0,000000	-1,636,014	1.97E-13
chr18	p11.31 - 11.23	3.954.575	8.055.875	0,000000	-0,786816	0,000000	1.40E-11
chr18	p11.21	11.286.577	13.885.315	0,000000	-0,362484	0,000000	9.21E-10
chr19	p13.3	5.843.077	6.851.877	0,412088	0,000000	0,000000	5.38E-32
chr19	q11 - q13.11	28.272.497	33.976.688	0,000000	-0,490308	0,000000	1.60E-109
chr21	p11.2	9.832.936	9.834.682	0,000000	-0,721174	0,000000	1.30E-36
chr21	q11.2 - q22.13	14.640.392	38.608.514	0,000000	-0,533194	0,000000	4,900e-324
chr21	q22.11 - 22.13	35.210.724	38.608.514	0,000000	-0,811744	0,000000	1.04E-19

Supplementary Table 10. Percentage of identity between the different modeled transcripts in the 20q and Xp loci. IP-EV: IP-sh scramble; IP-T: IP-shTRF2[1]

Chr.20q transcripts

Chr.Xp transcripts	IP-T	**4267.1**	**4268.1**	**4269.1**	**4270.1**	**4271.1**	**4272.1**	**4273.1**	**4274.1**	**4275.1**
	8745.1	39.3	39.1	53.9	46.3	50.9	36.3	45.1	46.9	37.9
	8743.1	37.4	41.2	58.6	38.2	36	47.2	44.2	41.2	40.3
	8744.1	40.1	36.4	55.9	35.1	30.4	35.5	35.8	42.1	33.7
	8744.2	37.4	36.4	55.9	45.2	45.4	35.5	35.8	42.1	39.5
	8746.1	41.4	39.8	38.2	42.5	49.8	38.6	42.4	50.3	39.1

Chr.20q transcripts

Chr.Xp transcripts	IP-EV	**4183.1**	**4184.1**	**4185.1**	**4186.1**	**4187.1**	**4188.1**	**4189.1**	**4190.1**	**4191.1**
	8070.1	40.9	48.9	39	42.2	38.6	41.4	41.8	43.8	40.7
	8074.1	36.5	41.4	40.4	37.1	43.2	82	42.3	52.8	39
	8073.1	37	38.4	44.5	36.1	42.6	71.6	45.1	34.4	45
	8071.1	43.5	38.4	44.5	36.1	42.6	50.8	39.7	43.5	45
	8072.1	37	47.8	40	38.2	43.1	71.6	45.1	34.4	38.4
	8072.2	38.6	47.8	40	38.2	43.1	71.6	45.1	34.4	38.4
	8075.1	32.8	51.1	37.2	48	54.8	42.4	46.6	41.6	38.2

Supplementary Table 11. Percentage of identity, size (bp) and presence of DNA repeats in the 20q-modeled transcripts with more than 50% homology with the Xp-modeled transcripts. IP-EV: IP-sh scramble; IP-T: IP-shTRF2[1]

	20q RNAs	**%Homology with Xp**	**Transcript size**	**Repeats**
IP-T	**4269.1**	53-55%	72 bp	LTR
	4271.1	50%	546 bp	SINE/LINE
	4274.1	50%	167 bp	-

IP-EV	**4188.1**	82-71%	90 bp	SINE/LINE
	4187.1	54%	185 bp	LTR
	4184.1	51%	219 bp	LINE/LTR

Supplementary Table 12. Percentage of Identity between the one-exon modeled transcript from the 20q and Xp loci. IP-EV: IP-sh scramble; IP-T: IP-shTRF2[1]

	20q-RNAs vs Xp-RNAs
IP-EV	37.70%
IP-T	39.20%

Supplementary Table 13. Primers for RNA-FISH and Northern blot probes

Region	**Primer**	**Sequence (5´-3´)**
TAR1	F-ProbeTAR1	GCAGAGTTCTTCTCAGGTCA
	R-ProbeTAR1	TGTCATGTGTGCATTAGGAA
	F-T7ProbeTAR1	ccaagcttctaatacgactcactatagggagaGCAGAGTTCTTCTCAGGTCA
	R-T7ProbeTAR1	ccaagcttctaatacgactcactatagggagaTGTCATGTGTGCATTAGGAA
DDX11L	F-ProbeDDX11L	TGTGTGGAAGTTCACTCCTG
	R-ProbeDDX11L	CAGGATGGAAGACAGATTGG
	F-T7ProbeDDX11L	ccaagcttctaatacgactcactatagggagaTGTGTGGAAGTTCACTCCTG
	R-T7ProbeDDX11L	ccaagcttctaatacgactcactatagggagaCAGGATGGAAGACAGATTGG
WASH	F-ProbeWASH	GCCTCTTAAGAACACAGTGG
	R-ProbeWASH	ATCTCTGGGAAAGGACCTG
	F-T7ProbeWASH	ccaagcttctaatacgactcactatagggagaGCCTCTTAAGAACACAGTGG
	R-T7ProbeWASH	ccaagcttctaatacgactcactatagggagaATCTCTGGGAAAGGACCTG
Chr20q	F-ProbeChr20q	CATTGCTGGTGGAAGACAG
	R-ProbeChr20q	GGACCCTCTGTAACAAATGAC
	F-T7ProbeChr20q	ccaagcttctaatacgactcactatagggagaCATTGCTGGTGGAAGACAG
	R-T7ProbeChr20q	ccaagcttctaatacgactcactatagggagaGGACCCTCTGTAACAAATGAC
ChrXpYp	F-ProbeChrXpYp	AACCTGCCGTATCTCACC
	R-ProbeChrXpYp	ACACTCCTGTAGTCCCACCT
	F-T7ProbeChrXpYp	ccaagcttctaatacgactcactatagggagaAACCTGCCGTATCTCACC
	R-T7ProbeChrXpYp	ccaagcttctaatacgactcactatagggagaACACTCCTGTAGTCCCACCT

Supplementary Table 14. Primers for promoter region cloning

Region	**Primer**	**Sequence (5´-3´)**
20qPr1	F-0qPr1	CGCGGTACCAAGCCTAAGTGTGGCAAGTC
	R-0qPr1	CCCGCTCGAGCAAGGTGTGTGCTGAGAAAG
20qPr2	F-0qPr2	CGCGGTACCGGAATCATGAAGTATGTGACC

	R-0qPr2	CCCGCTCGAGGGCCTAGTATCCGGAATATACA
20qPr3	F-0qPr3	CGCGGTACCTGTGCGTGTTTCTGTGTC
	R-0qPr3	CCCGCTCGAGGTTCAAGGCCCTCCACAG
20qPr4	F-0qPr4	CGCGGTACCGTGGAGTCAGACCACACG
	R-0qPr4	CCCGCTCGAGTACCAGAAGCTGGAGAAAGG

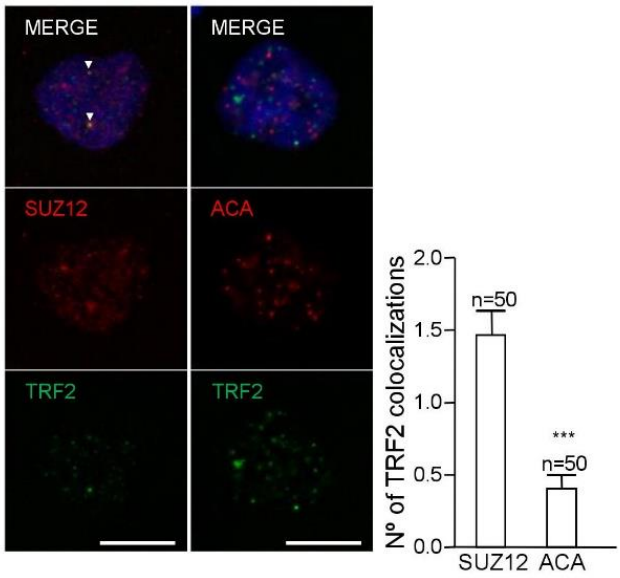
Supplementary Table 15. Primers for the amplification of DNA regions with predicted off-target mutations

gRNA	Primer	Sequence (5´-3´)
gRNA Start2	F-gS2Off1	TACAAGGTCCGGAGAAGC
	R-gS2Off1	GAGCGGAGGAGGGAGGAC
	F-gS2Off2	TGAGTACCAAGGGTCACCTC
	R-gS2Off2	GGACGGCTGGTTCTCAGG
	F-gS2Off3	GGATGATCTGGGAGAAGC
	R-gS2Off3	GGGAATCTGCAGGACAAAC
	F-gS2Off4	GGGCAGGTAAGGACGTTT
	R-gS2Off4	GCTCCTCACTCCCAAATC
gRNA End1	F-gE1Off1	CTCTAGTTCATGCAGGATGC
	R-gE1Off1	GAGTGAGAAGAGAGCCTTGG
	F-gE1Off2	AATCCAGGAGGCAGAGGT
	R-gE1Off2	CAGGTCAGGGAGTCAACAAC
	F-gE1Off3	TAGGTCACAGAGCAGTAGGG
	R-gE1Off3	GTCCCTAAGGCATTTAGCA
	F-gE1Off4	TTTCCCAACTATTTGCTGAT
	R-gE1Off4	GTTCTGCACCTTATAAACCAG

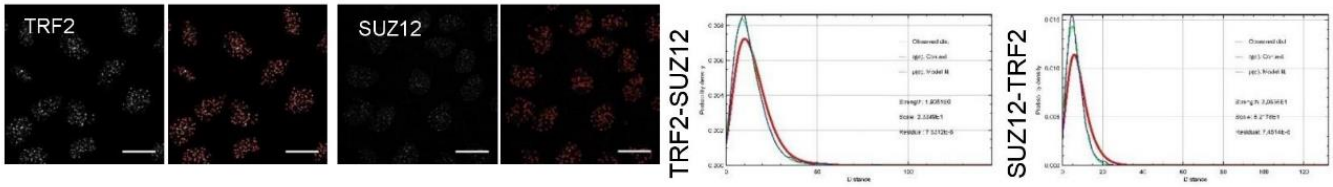
Supplementary References

1. Porro, A. *et al.* Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat Commun* **5**, 5379 (2014).

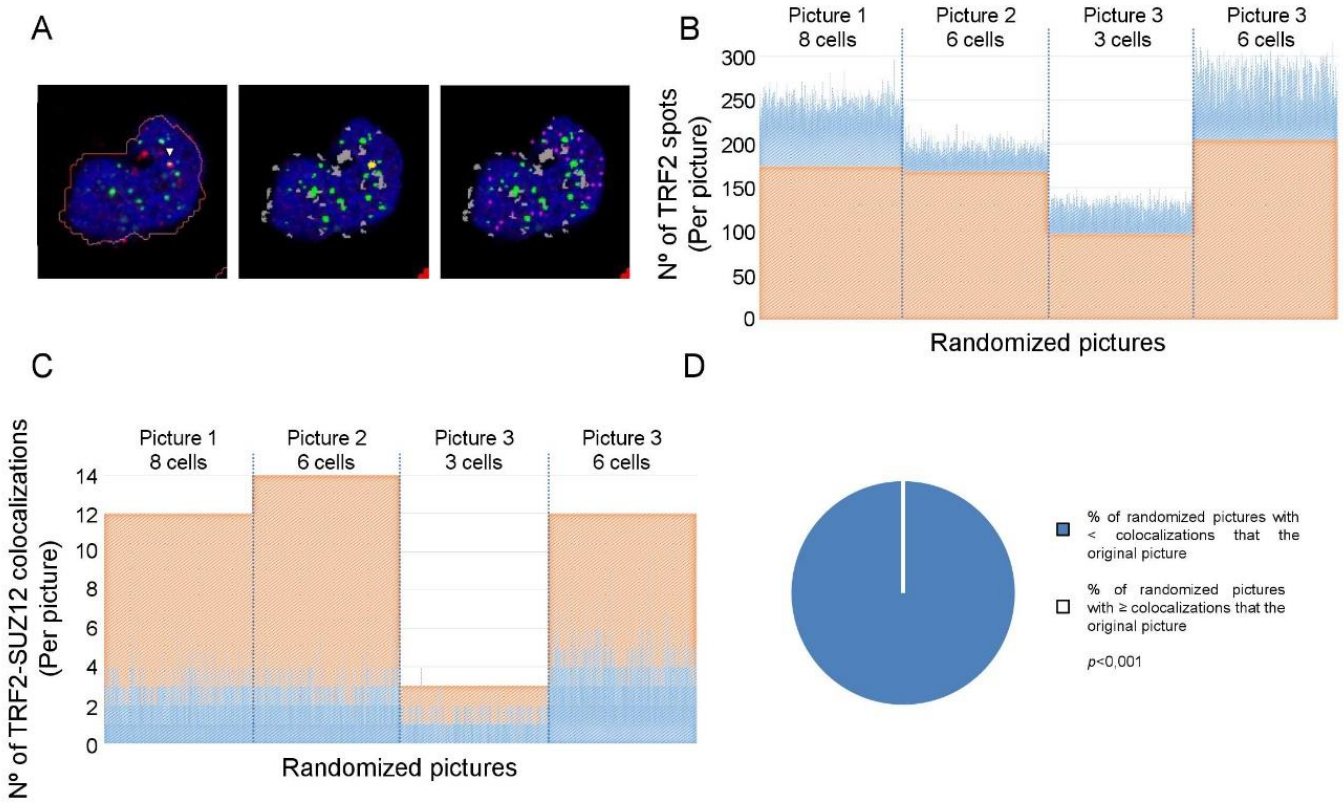
Anexo 1.2: TERRA recruitment of polycomb to telomeres is essential for histone trymethylation marks at telomeric heterochromatin



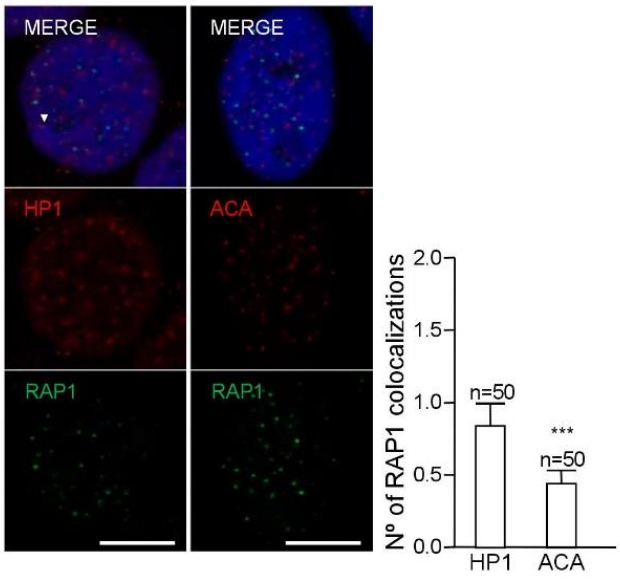
Supplementary Figure 1. Quantification of the random colocalization between the centromere and the telomeric protein TRF2. Representative images of the average number of colocalizations found on double inmunostaining to detect the telomere protein TRF2 (green) and the PRC2 protein SUZ12 (red) (Left panel) or between TRF2 (green) and the centromere using the Anti-Centromere Antibody (ACA) (red) (Right panel) in U2OS cells. (*Graph*) Quantification of the co-localization between TRF2 and SUZ12 or TRF2 and ACA. Arrowheads indicate co-localization events. Scale bar, 10 μm. Student's t-test was used for the statistical analysis (*p <0.05, ** p <0.01 and *** p <0.001).



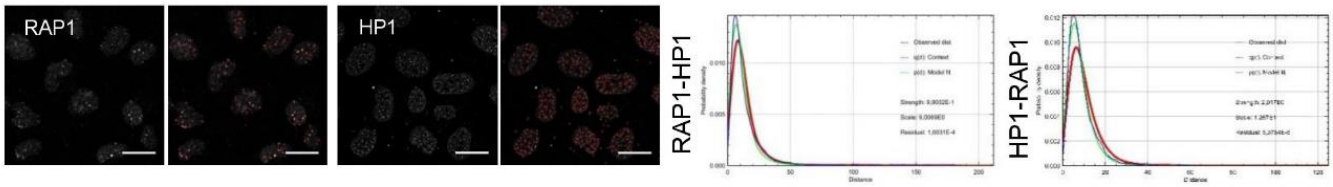
Supplementary Figure 2. MosaicIA radomization approach to prove the interaction of SUZ12 with the telomere. Representative images of object identification using MosaicIA interaction plugin of Fiji for TRF2 and SUZ12 signal. Red circles: Identified objects. Scale: 20*μM*. (*Graphs*) Graphical representation of the comparison of the interacting potential between SUZ12 and TRF2. Blue line: distribution of NND between objects identified in TRF2 images vs SUZ12 images. Green line: modelization of interaction potential using a Plummer potential. Red line: probability density function resulting of the calculation of NND between TRF2 and SUZ12 images if the objects in TRF2 were distributed randomly, in a completely independent manner of the distribution of SUZ12 (l*eft graph*). Similarly, it was also calculated for SUZ12 objects distributed randomly, in a completely independent manner of the distribution of TRF2 (*right graph*). Strength is a measure of the degree of dependence between TRF2 and SUZ12 objects distribution. The strength is superior to zero, indicating that the spatial distribution of TRF2 is dependent upon the spatial distribution of SUZ12. When compared against 10000 Monte Carlo samples of point distributions corresponding to the null hypothesis of “no interaction”, the results are statistically significant (p<0,001).



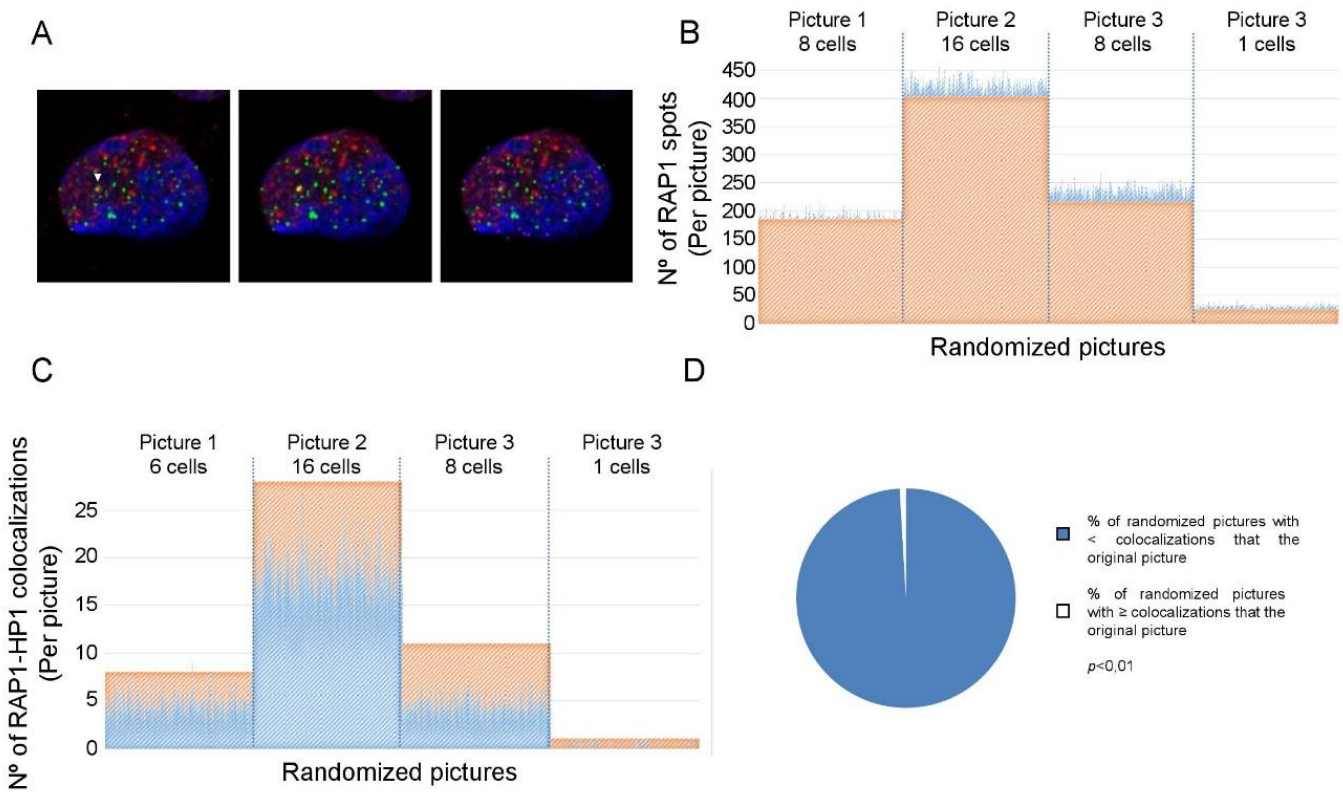
Supplementary Figure 3. Definiens radomization approach to prove the interaction of SUZ12 with the telomere. (A) Representative images of the co-localizations found on double inmunostaining to detect the telomere protein TRF2 (green) and the PRC2 protein SUZ12 (red). (*Left*) Nuclear identification by the Definiens Developer XD.2 software(red line). Arrowheads indicate real co-localization events. (*Middle*) Representative example of the TRF2 spots identification by the Definiens Developer XD.2 program (green). Co-localization events identified by the program is shown in yellow. The gray areas represent areas of low DNA signal that were not used for the analysis. (*Right*) Representative example of a randomized image generated by the Definiens Developer XD.2 program. Virtual TRF2 spots are shown (magenta). **(B)** Graphical representation of the total real and virtual TRF2 spots identified by the Definiens Developer XD.2 program in four different pictures (each picture contains different number of nuclei). The orange bars represents the number of real TRF2 spots in the originals pictures and the blue bars represent the virtual TRF2 spots generated by the program in every random permutation. **(C)** Graphical representation of the total real and virtual TRF2-SUZ12 co-localizations events identified by the Definiens Developer XD.2 program in four different pictures. The orange bars represents the number of TRF2-SUZ12 real co-localizations identified in the originals pictures and the blue bars represent the virtual TRF2-SUZ12 co-localizations identified in every random permutation. **(D)** Diagram showing the % of randomized pictures with < co-localizations that the original picture versus the % of randomized pictures with ≥ co-localizations that the original picture. The result show that the co-localization events between SUZ12 and TRF2 are not due to random colocalization and this is statistically significant (p<0,001). The p-value is calculated as the probability of finding a randomized picture with ≥ co-localizations than in the original picture. This is calculated with the number of randomized pictures with ≥ co-localizations than in the original picture divided by the total number of randomized pictures.



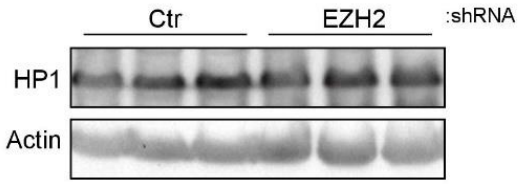
Supplementary Figure 4. Quantification of the random colocalization between the centromere and the telomeric protein RAP1. Representative images of the average number of colocalizations found on double inmunostaining to detect the telomere protein RAP1 (green) and the HP1 protein (red) (Left panel) or between RAP1 (green) and (ACA) (red) (Right panel) in U2OS cells. (*Graph*) Quantification of the co-localization between RAP1 and HP1 or RAP1 and ACA. Arrowheads indicate co-localization events. Scale bar, 10 μm. Student's t-test was used for the statistical analysis (*p <0.05, ** p <0.01 and *** p <0.001).



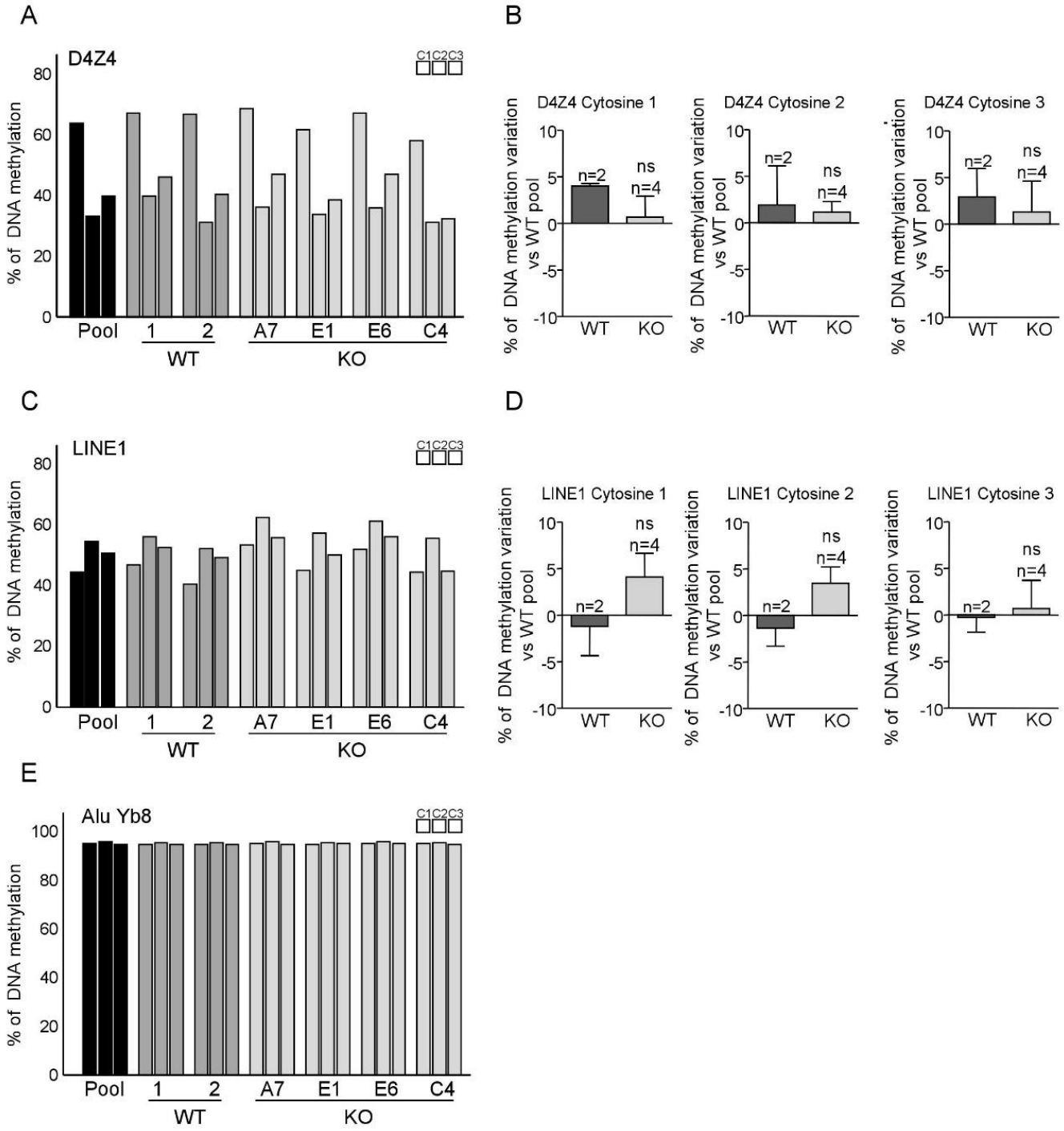
Supplementary Figure 5. MosaicIA radomization approach to prove the interaction of HP1 with the telomere. Representative images of object identification using MosaicIA interaction plugin of Fiji for RAP1 and HP1 signal. Red circles: Identified objects. Scale: 20*μM*. (*Graphs*) Graphical representation of the comparison of the interacting potential between HP1 and RAP1. Blue line: distribution of NND between objects identified in RAP1 images vs HP1 images. Green line: modelization of interaction potential using a Plummer potential. Red line: probability density function resulting of the calculation of NND between RAP1 and HP1 images if the objects in RAP1 were distributed randomly, in a completely independent manner of the distribution of HP1(*left graph*). Similarly, it was also calculated for HP1 objects distributed randomly, in a completely independent manner of the distribution of RAP1. Strength is a measure of the degree of dependence between RAP1 and HP1 objects distribution. The strength is superior to zero, indicating that the spatial distribution of RAP1 is dependent upon the spatial distribution of RAP1. When compared against 10000 Monte Carlo samples of point distributions corresponding to the null hypothesis of "no interaction", the results showed are statistically significant (p<0,001).



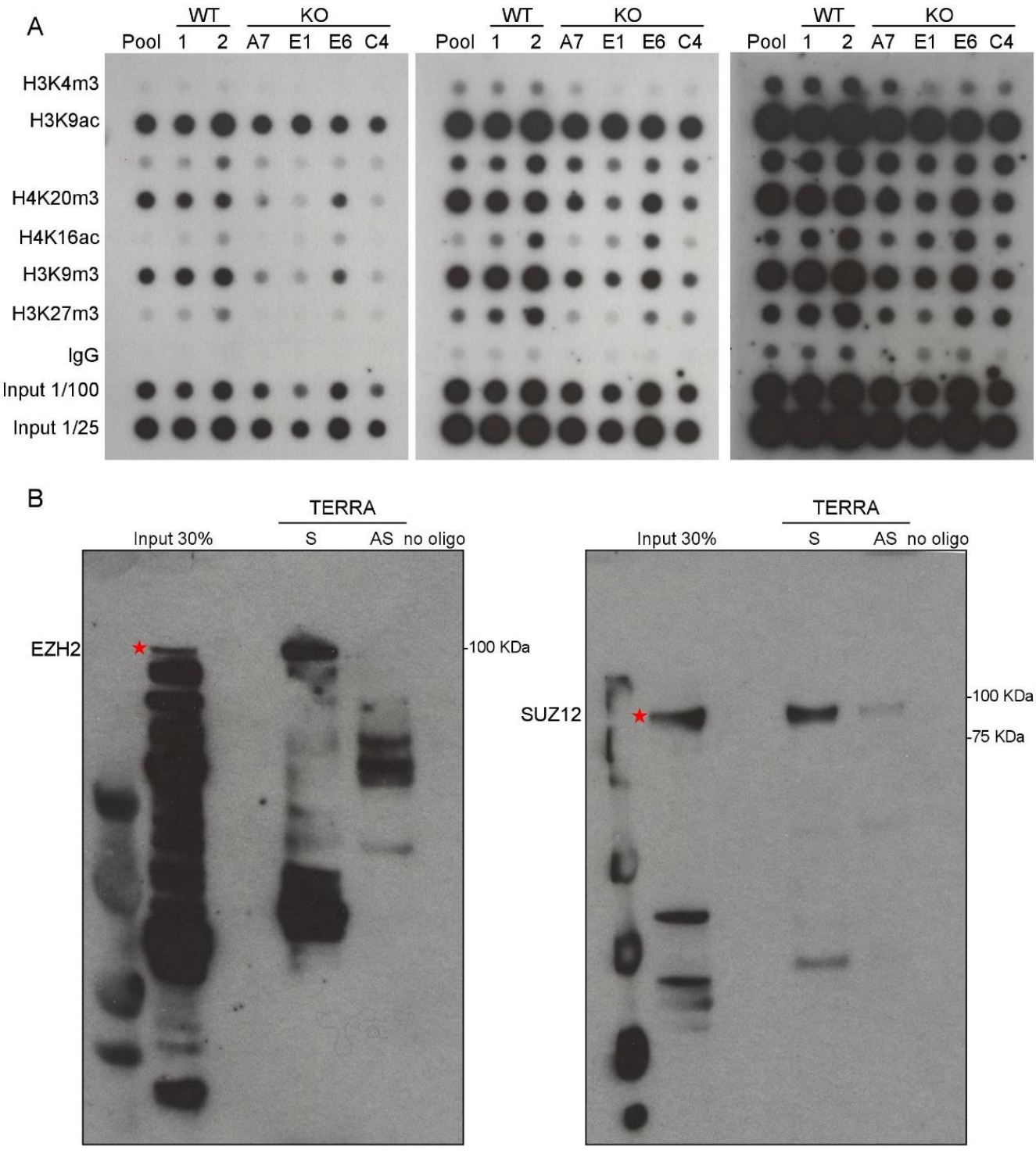
Supplementary Figure 6. Definiens radomization approach to prove the interaction of HP1 with the telomere. (A) Representative images of the co-localizations found on double inmunostaining to detect the telomere protein RAP1 (green) and the HP1 protein (red). (*Left*). Arrowheads indicate real co-localization events. (*Middle*) Representative example of the RAP1 spots identification by the Definiens Developer XD.2 program (green). Co-localization events identified by the program is shown in yellow. (*Right*) Representative example of a randomized image generated by the Definiens Developer XD.2 program. Virtual RAP1 spots are shown (magenta). (**B**) Graphical representation of the total real and virtual RAP1 spots identified by the Definiens Developer XD.2 program in four different pictures (each picture contains different number of nuclei). The orange bars represents the number of real RAP1 spots in the originals pictures and the blue bars represent the virtual RAP1 spots generated by the program in every random permutation. **(C)** Graphical representation of the total real and virtual RAP1-HP1 co-localizations events identified by the Definiens Developer XD.2 program in four different pictures. The orange bars represents the number of RAP1-HP1 real co-localizations identified in the originals pictures and the blue bars represent the virtual RAP1-HP1 co-localizations identified in every random permutation. **(D)** Diagram showing the % of randomized pictures with < co-localizations that the original picture versus the % of randomized pictures with ≥ co-localizations that the original picture. The result show that the co-localization events between HP1 and RAP1 are not due to random colocalization and this is statistically significant (p<0,01). The p-value is calculated as the probability of finding a randomized picture with ≥ co-localizations than in the original picture. This is calculated with the number of randomized pictures with ≥ co-localizations than in the original picture divided by the total number of randomized pictures.

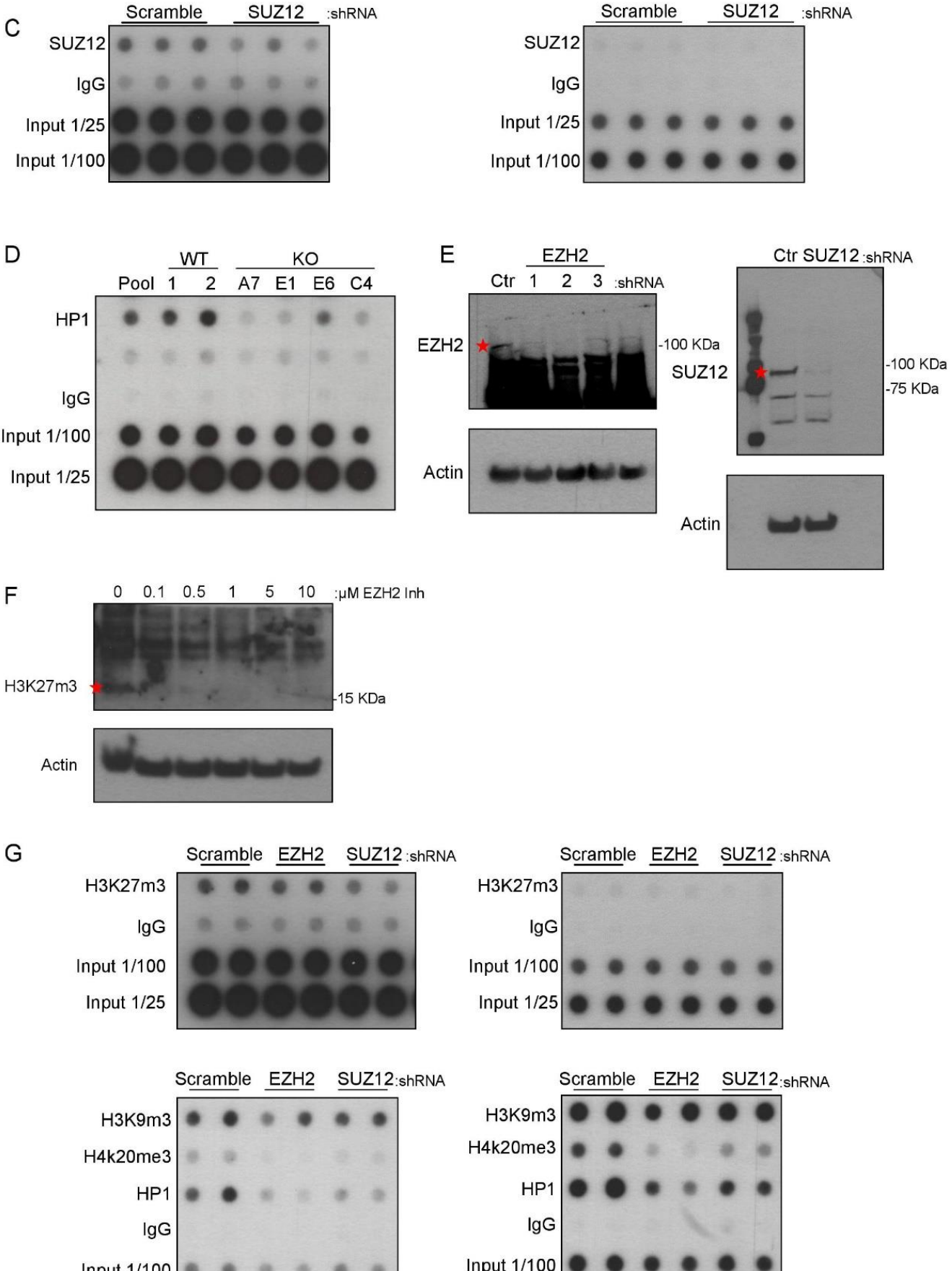


Supplementary Figure 7. HP1 levels do not change upon EZH2 donwregulation. U2OS cells were infected with a lentivirus encoding a shRNA against EZH2. After puromycin selection, total protein extracts were obtained and use for western blot detection of HP1. Actin was used as loading control.


A
D4Z4
C1C2C3
% of DNA methylation
Pool
1
2
A7
E1
E6
C4
WT
KO
B
D4Z4 Cytosine 1
D4Z4 Cytosine 2
D4Z4 Cytosine 3
% of DNA methylation variation vs WT pool
n=2
ns
n=4
C
LINE1
D
LINE1 Cytosine 1
LINE1 Cytosine 2
LINE1 Cytosine 3
E
Alu Yb8

Supplementary Figure 8. TERRAs do not change subtelomeric and global DNA methylation. (A) Graph showing the percentage of methylation of the subtelomeric repeat D4Z4 by bisulfite pyrosequencing at the first three cytosines (C1, C2 and C3) in the U2OS WT pool, WT clones (#1 and 2) and in the 20q-TERRA KO clones (#A7, E1, E6, and C4). **(B)** Percentage of DNA methylation variation in the subtelomeric repeat D4Z4 respect to the WT pool in the WT clones and in the 20q-TERRA KO (mean values±s.e.m., n=independent clones). **(C)** Graph showing the percentage of DNA methylation of the Line1 repeats by bisulfite pyrosequencing at the first three cytosines in the U2OS WT pool, WT clones (#1 and 2) and in the 20q-TERRA KO clones (#A7, E1, E6, and C4). **(D)** Percentage of DNA methylation variation in the Line1 repeat respect to the WT pool in the WT clones and 20q-TERRA KO clones (mean values±s.e.m., n=independent clones). **(E)** Percentage of methylation of Alu repeat by bisulfite pyrosequencing at the first threet cytosines in the U2OS WT pool, WT clones (#1 and 2) and in the 20q-TERRA KO clones (#A7, E1, E6, and C4). The methylation at Alu repeats are use as control. Student's t-test was used for the statistical analysis (*p<0.05, ** p <0.01 and *** p <0.001).


A
WT
KO
Pool
1
2
A7
E1
E6
C4
H3K4m3
H3K9ac
H4K20m3
H4K16ac
H3K9m3
H3K27m3
IgG
Input 1/100
Input 1/25
B
TERRA
Input 30%
S
AS
no oligo
EZH2
100 KDa
SUZ12
100 KDa
75 KDa


C
Scramble
SUZ12
:shRNA
SUZ12
IgG
Input 1/25
Input 1/100
Scramble
SUZ12
:shRNA
SUZ12
IgG
Input 1/25
Input 1/100
D
WT
KO
Pool 1 2 A7 E1 E6 C4
HP1
IgG
Input 1/100
Input 1/25
E
EZH2
Ctr 1 2 3 :shRNA
EZH2
-100 KDa
Actin
Ctr SUZ12 :shRNA
SUZ12
-100 KDa
-75 KDa
Actin
F
0 0.1 0.5 1 5 10 :μM EZH2 Inh
H3K27m3
-15 KDa
Actin
G
Scramble EZH2 SUZ12 :shRNA
H3K27m3
IgG
Input 1/100
Input 1/25
Scramble EZH2 SUZ12 :shRNA
H3K27m3
IgG
Input 1/100
Input 1/25
Scramble EZH2 SUZ12 :shRNA
H3K9m3
H4k20me3
HP1
IgG
Input 1/100
Input 1/25
Scramble EZH2 SUZ12 :shRNA
H3K9m3
H4k20me3
HP1
IgG
Input 1/100
Input 1/25

Supplementary Figure 9. Full Blots and Gels. (A) Full ChIP Dot-Blot from the main figure 3. Different exposures are shown. Only relevant lanes are labeled. **(B)** Full Western Blot from the main figure 4A. **(C)** Full ChIP Dot-Blot from the main figure 4C. Different exposures are shown. **(D)** Full ChIP Dot-Blot from the main figure 4E. Only relevant lanes are labeled. **(E)** Full Western Blot from the main figure 5A. **(F)** Full Western Blot from the main figure 5B. **(G)** Full ChIP Dot-Blot from the main figure 5D and 5F. Different exposures are shown. In the Western Blots red stars indicated the bands shown in the main figure.

Supplementary Table 1. shRNAs and primers

.

Primer	Sequence (5´-3´)
shRNA scramble	CCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGG
shRNA EZH2 1	GCTAGGTTAATTGGGACCAAACTCGAGTTTGGTCCCAATTAACCTAG
shRNA EZH2 2	CCAACACAAGTCATCCCATTACTCGAGTAATGGGATGACTTGTGTTGG
shRNA EZH2 3	CCCAACATAGATGGACCAAATCTCGAGATTTGGTCCATCTATGTTGGG
shRNA SUZ12	GCTGACAATCAAATGAATCATCTCGAGATGATTCATTTGATTGTCAGC
F-U620qE1Not	GCGAGcggccgcacatgtgagggcctatttcc
R-U620qE1Not	CGGTGcggccgcAtggggagagtgaagcagaa

Anexo 1.3: TERRA regulate the transcriptional landscape of pluripotent cells throught TRF1-dependent recruitment of PRC2

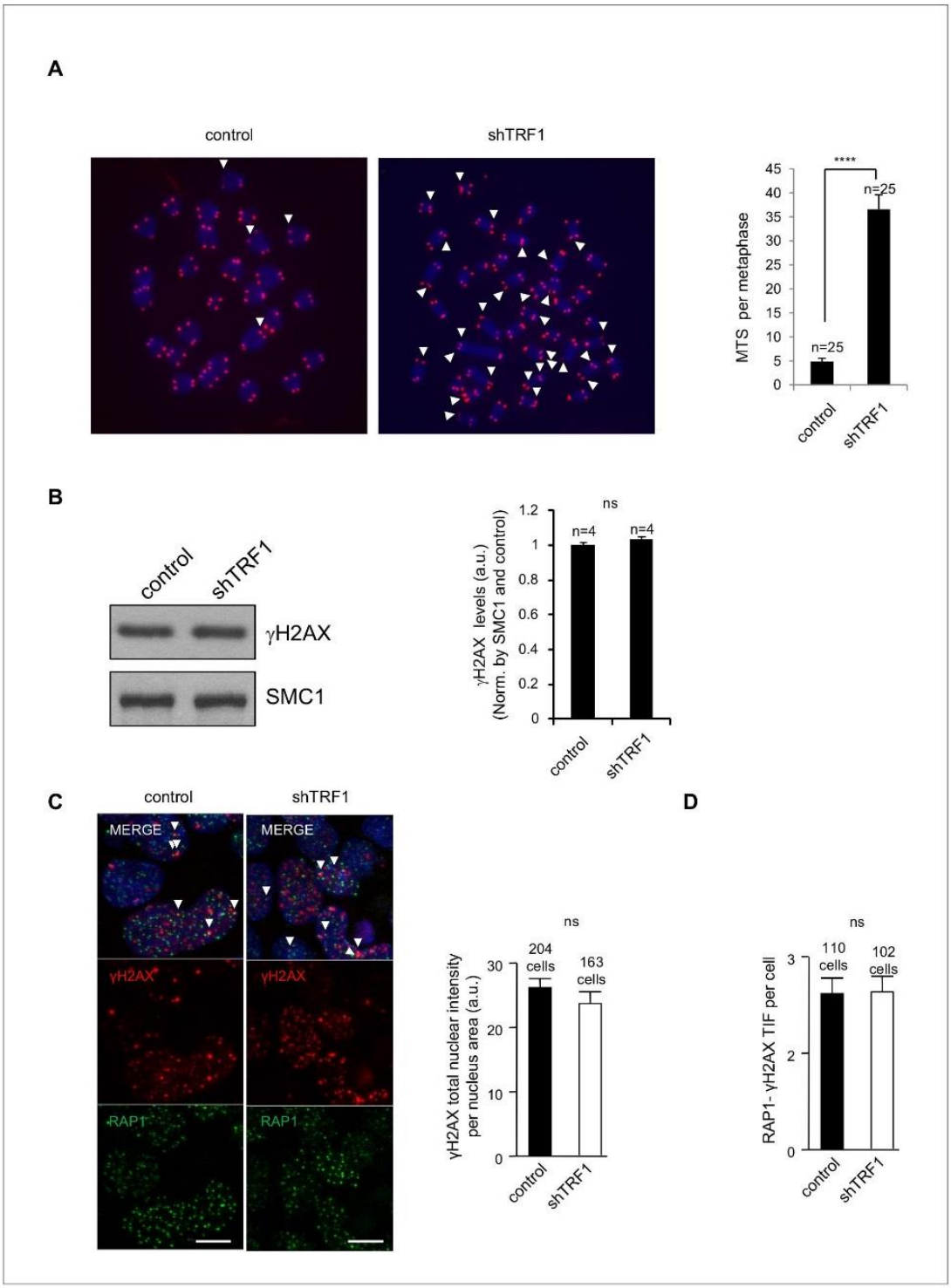


Figure 1—figure supplement 1. TRF1 abrogation in 2i-grown iPS cells does not increase DNA damage. (**A**) Left, representative telomeric FISH imaging of metaphase-stage control and TRF1-depleted *Trp53*$^{-/-}$ 2i-grown iPS cells. White arrows point to the presence of multitelomeric signals (MTS).
Figure 1—figure supplement 1 continued on next page

Figure 1—figure supplement 1 continued

Right, quantification of the number of MTS per metaphase. Note the dramatic increase in MTS upon TRF1 depletion. N = number of metaphases. Error bars = standard errors of the means (SE). Statistical analysis was carried out using Student's t-test. The experiment was performed once. (**B**) (Left) Representative image of Western blots of γH2AX and SMC1 (as loading control) from control and TRF1-depleted iPS cells. (Right) Quantification of γH2AX levels in four independent experiments. Note that γH2AX levels are not affected by the depletion of TRF1 protein. n = number of independent experiments. Error bars = SE. Statistical analysis, Student's t-test. (**C**) (Left) Representative double-immunofluorescence imaging of telomeric protein RAP1 and the DNA-damage marker γH2AX in control iPS and TRF1-depleted iPS. (Right) Quantification of γH2AX intensity per nucleus. Note that there is no increase of DNA damage upon TRF1 abrogation. Data were obtained from one experiment. n = number of cells analyzed from each sample. Error bars = SE. Statistical analysis, Student's t-test. (**D**) Quantification of the number of TIFs (telomere induced foci) present in the cells shown in (**C**), measured as the number of colocalizations of the telomeric protein RAP1 with the DNA damage marker γH2AX. Note that there is no increase of telomeric damage upon TRF1 deletion in these conditions. Data were obtained from one experiment. n = number of cells analyzed from each sample. Error bars = SE. Statistical analysis, Student's t-test.

DOI: https://doi.org/10.7554/eLife.44656.003

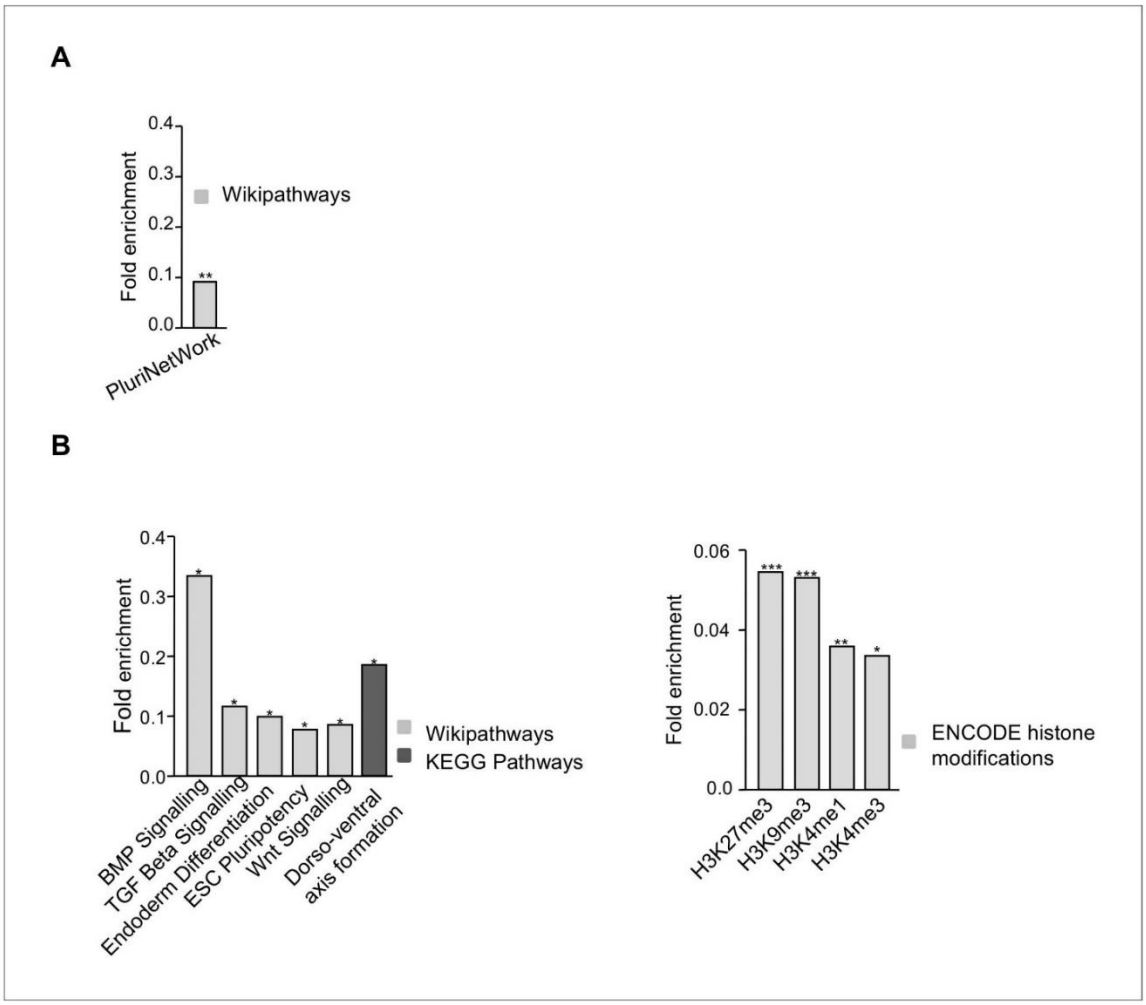


Figure 1—figure supplement 2. Functional annotation of genes altered when depleting TRF1. (**A**) Genes that are downregulated upon TRF1 deletion are involved in signaling pathways that regulate the pluripotency of stem cells. (**B**) Genes that are upregulated upon TRF1 deletion were enriched in important pathways regulating pluripotency and differentiation (left) and in several epigenetics marks (right), including the polycomb-related chromatin marks H3K27me3 and H3K4me3.

DOI: https://doi.org/10.7554/eLife.44656.004

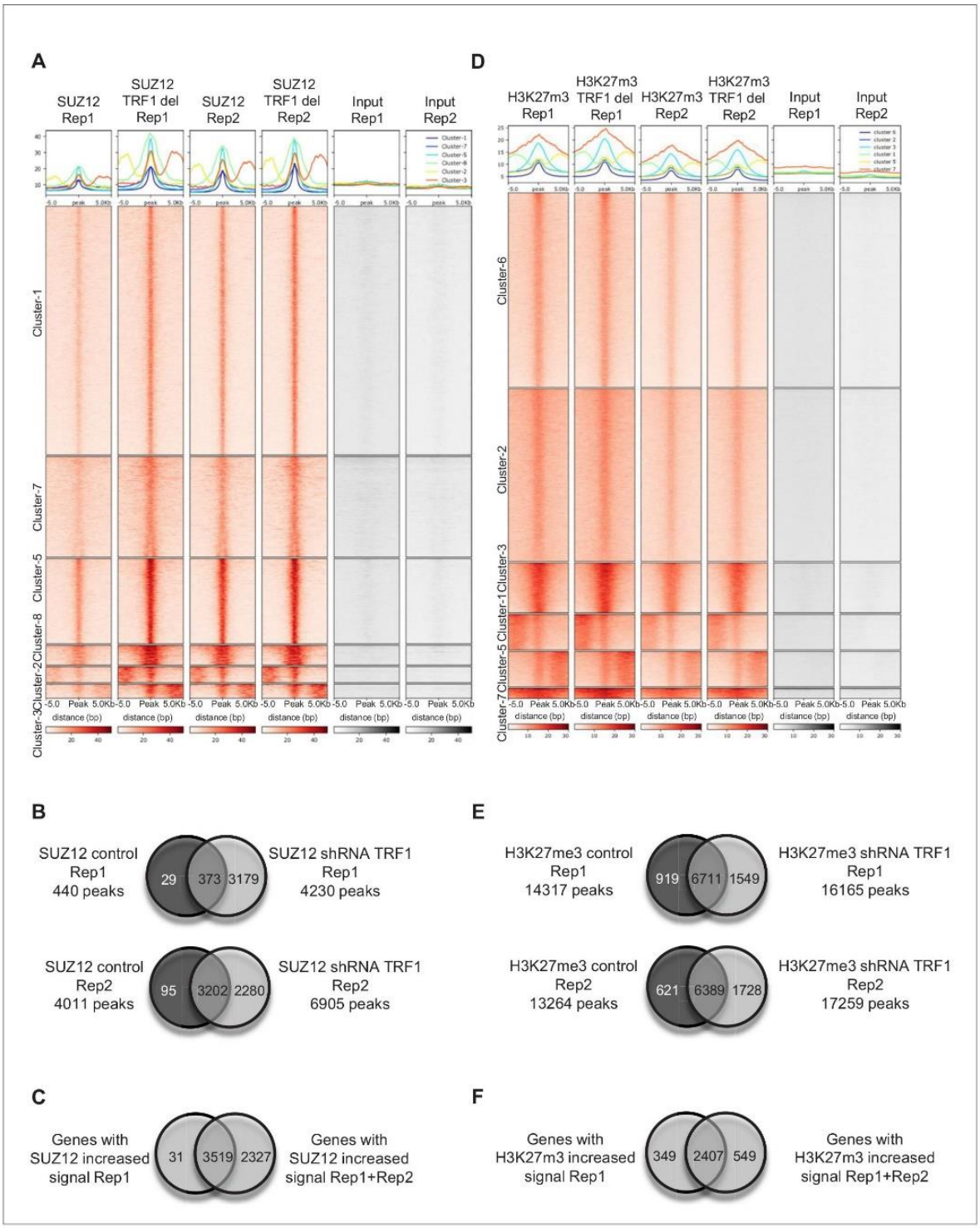


Figure 2—figure supplement 1. Validation of SUZ12 and H3K27me3 ChIP-seq. (**A**) Heatmaps of the two biological replicates of SUZ12 ChIP-seq, with reads plotted around the peaks for both replicates. Note that the pattern of both replicates is the same. All of the clusters include regions where SUZ12

Figure 2—figure supplement 1 continued on next page

Figure 2—figure supplement 1 continued

increases upon TRF1 depletion in both replicates. (**B**) Venn diagrams showing the overlapping of genes associated with peaks in control samples and samples in which TRF1 have been depleted for both SUZ12 ChIP-seq replicate 1 (top) and SUZ12 ChIP-seq replicate 2 (bottom). (**C**) Venn diagram showing the overlapping of genes that showed increased SUZ12 mark upon TRF1 depletion in replicate 1 of SUZ12 ChIP seq (Clusters 1, 2 and 3 of ***Figure 2B***) with genes that showed increased SUZ12 mark upon TRF1 depletion in both SUZ12 ChIP seq replicates (all of the clusters of (**A**)). Note that that 99% of the genes that show increased levels of SUZ12 upon TRF1 depletion in the first experiment also show higher levels of SUZ12 in the second replicate, confirming the high similarity between the two ChIP-seq replicates. (**D**) Heatmaps of the two biological replicates of H3K27me3 ChIP-seq, with reads plotted around the peaks for both replicates. Note that the pattern is the same for both replicates. Clusters 3 and 7 include regions where H3K27me3 increases upon TRF1 depletion in both replicates. (**E**) Venn diagrams show the overlapping of genes associated with peaks in control samples and in samples in which TRF1 has been depleted for both H3K27me3 ChIP-seq replicate 1 (top) and H3K27me3 ChIP-seq replicate 2 (bottom). (**F**) Venn diagram showing the overlapping of genes that showed increased H3K27me3 mark upon TRF1 depletion in H3K27me3 ChIP seq replicate 1 (Cluster 1 of ***Figure 2F***) with genes that showed increased H3K27me3 mark upon TRF1 depletion in both H3K27me3 ChIP seq experiments (Clusters 3 and 7 of (**D**)). Note that that 90% of the genes that show increased levels of H3K27me3 upon TRF1 depletion in the first experiment also show higher levels of H3K27me3 in the second replicate, confirming the close similarity between the two ChIP-seq experiments.

DOI: https://doi.org/10.7554/eLife.44656.006

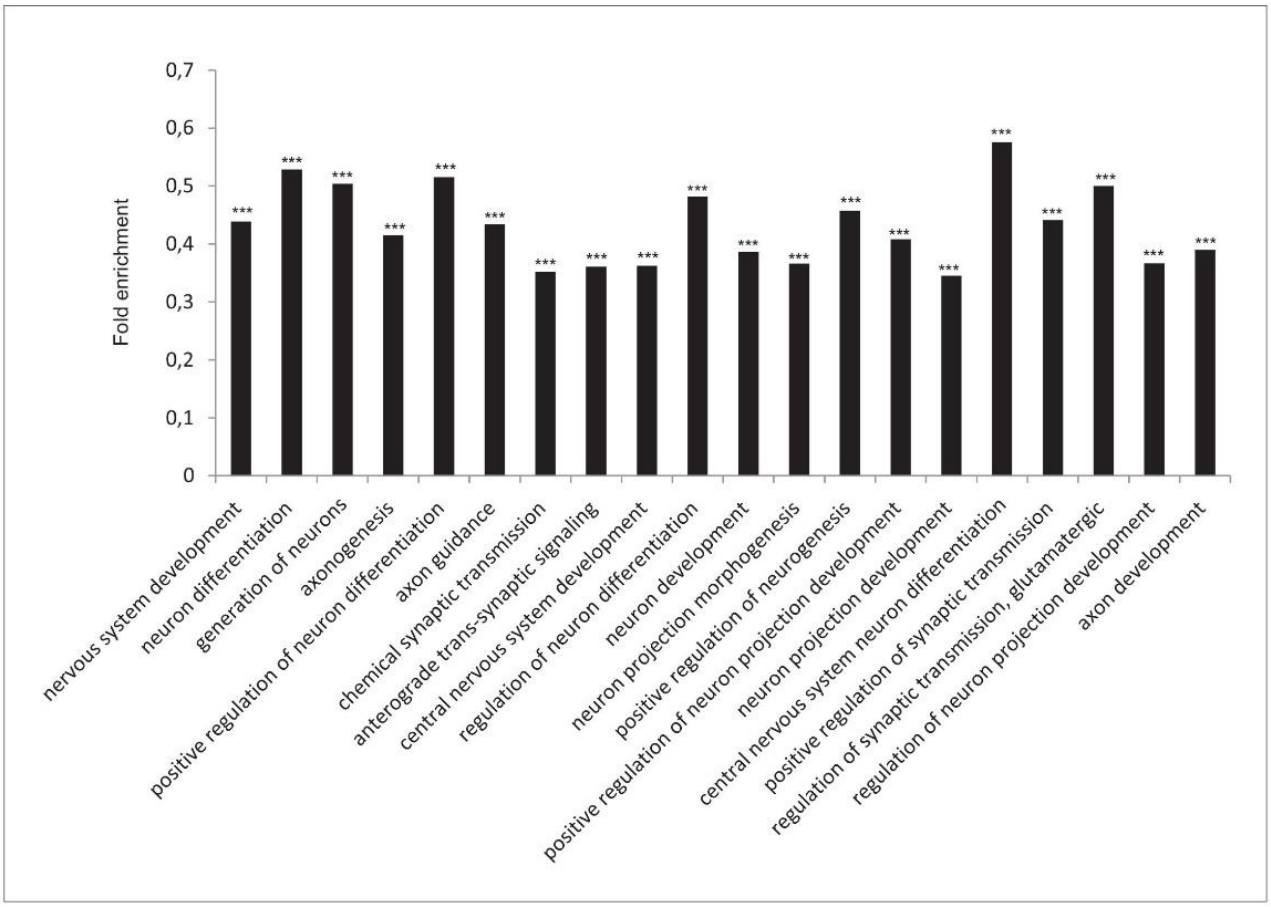


Figure 2—figure supplement 2. Functional annotation analysis of genes enriched in SUZ12 upon TRF1 depletion. (**A**) GO Biological process analysis of the genes associated with the genome sites were increased binding of SUZ12 protein was observed upon depletion of TRF1. Genes annotated in Clusters 1, 2 and 3 of the SUZ12 heatmap (***Figure 2B***) were used. Note that these genes are highly involved in processes related to the nervous system. *** = Adjusted p-value <0.0001.

DOI: https://doi.org/10.7554/eLife.44656.007

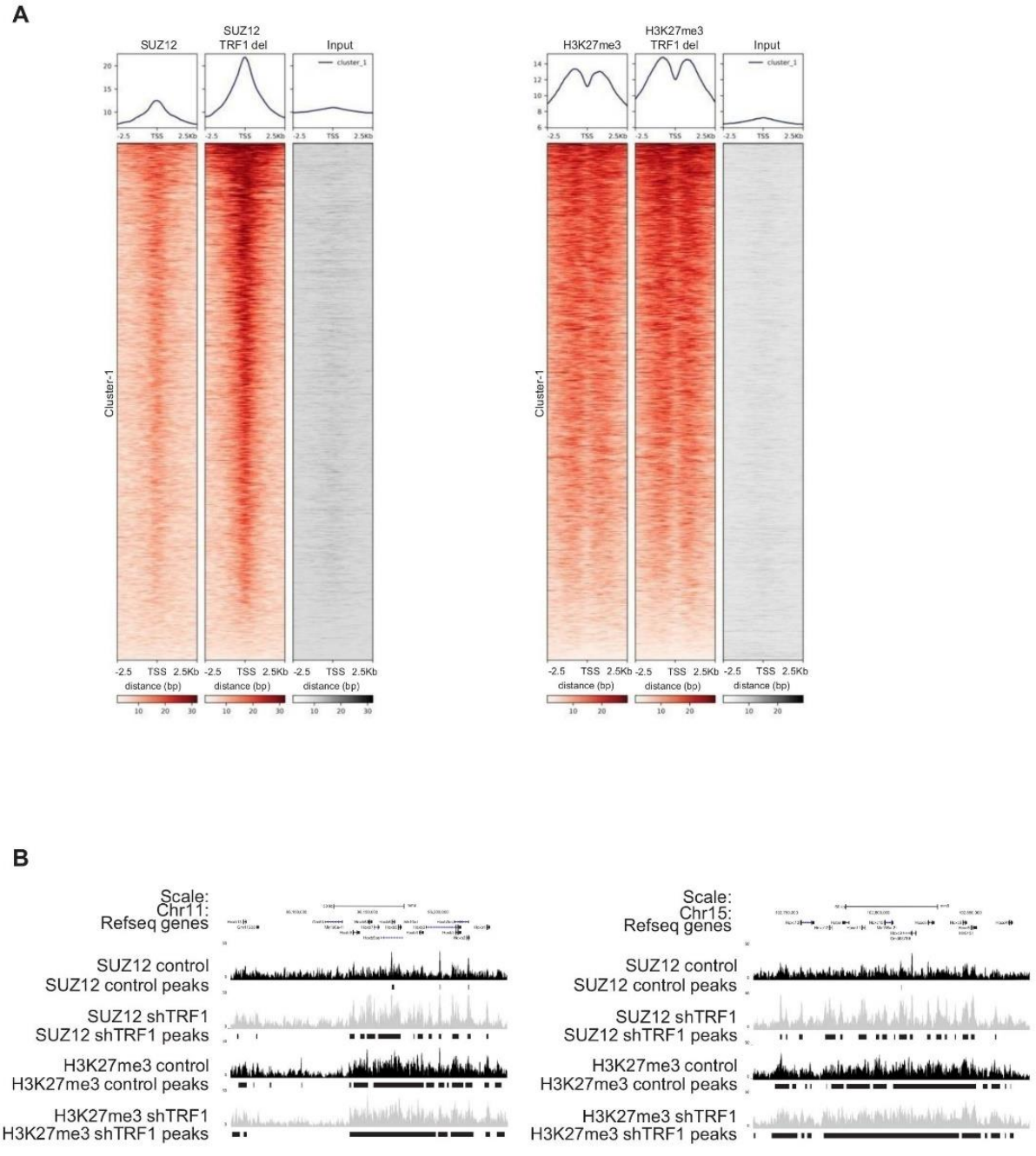
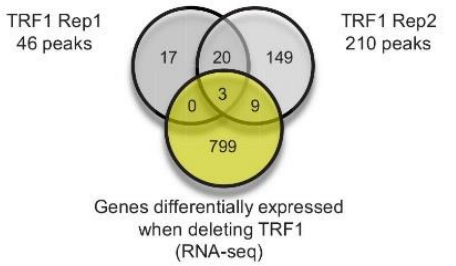


Figure 2—figure supplement 3. Abrogation of TRF1 in 2i-grown iPS cells induces the loss of the naïve state and a transition to a primed or differentiated state. (**A**) Heatmaps of the reads of SUZ12 and H3K27me3 around 2.5 Kb of the transcriptional start sequence (TSS) of bivalent genes. Note the dramatic increase of SUZ12 deposition at the TSS of bivalent genes. Increase in H3K27me3 at these sites was also clear but moderate. (**B**) Gain of SUZ12 and H3K27me3 peaks in some *Hox* clusters.

DOI: https://doi.org/10.7554/eLife.44656.008

A



B

Gene Name	Detailed Annotation
Gm14496	(TTAGGG)n|Simple_repeat|Simple_repeat
G530011O06Rik	(TTAGGG)n|Simple_repeat|Simple_repeat
Ptchd3	(TTAGGG)n|Simple_repeat|Simple_repeat
Adgre5	(CCCTAA)n|Simple_repeat|Simple_repeat
Rfc3	(TTAGGG)n|Simple_repeat|Simple_repeat
Morc3	(CCCTAA)n|Simple_repeat|Simple_repeat
4930559C10Rik	(TTAGGG)n|Simple_repeat|Simple_repeat
Gm3376	(CCCTAA)n|Simple_repeat|Simple_repeat
Creb3l1	(CCCTAA)n|Simple_repeat|Simple_repeat
Flt1	(TTAGGG)n|Simple_repeat|Simple_repeat
Fam43a	(TTAGGG)n|Simple_repeat|Simple_repeat
Nkx2-6	(CCCTAA)n|Simple_repeat|Simple_repeat
Ajap1	(TTAGGG)n|Simple_repeat|Simple_repeat
Lama1	(CCCTAA)n|Simple_repeat|Simple_repeat
4930474H20Rik	(CCCTAA)n|Simple_repeat|Simple_repeat
Ikzf2	(TTAGGG)n|Simple_repeat|Simple_repeat
Pde1c	(TTAGGG)n|Simple_repeat|Simple_repeat
Bbox1	(TTAGGG)n|Simple_repeat|Simple_repeat
Pitx2	(TTAGGG)n|Simple_repeat|Simple_repeat
Ppp1r9a	(CCCTAA)n|Simple_repeat|Simple_repeat
Olfr281	Intergenic
Cdk12	promoter-TSS (NM_026952)
Epm2a	Intergenic

C

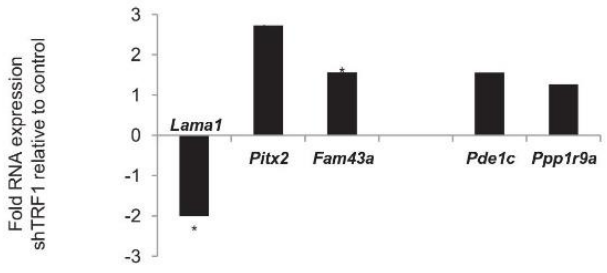


Figure 2—figure supplement 4. Extratelomeric binding of TRF1 in 2i-grown iPS cells. (**A**) Venn diagram showing the overlap of genes annotated to TRF1 peaks in both replicates of TRF1 ChIP-seq with the whole set of genes

Figure 2—figure supplement 4 continued on next page

Figure 2—figure supplement 4 continued

differentially expressed when depleting TRF1, as obtained by RNA-seq. Note that only three genes that are bound by TRF1 show significantly altered expression upon TRF1 depletion. (**B**) List of genes annotated to TRF1 peaks in both replicates of TRF1 ChIP-seq. Detailed annotation shows the binding sequence corresponding to each peak. Note that most of the peaks' binding sites correspond to telomeric repetitions (TTAGGG or CCCTAA). Gene targets of ZFP322A are labeled in bold. (**C**) Changes in the expression of gene targets of TRF1 (*Lama1*, *Pitx2* and *Fam43a*) and of targets of ZFP322A (*Lama1*, *Pde1c* and *Ppp1r9a*) that are significantly altered when depleting TRF1, as determined by RNA-seq. Note that two of the gene targets of ZFP322A were upregulated in the RNA-seq, whereas one of them, *Lama1*, was significantly downregulated.

DOI: https://doi.org/10.7554/eLife.44656.009

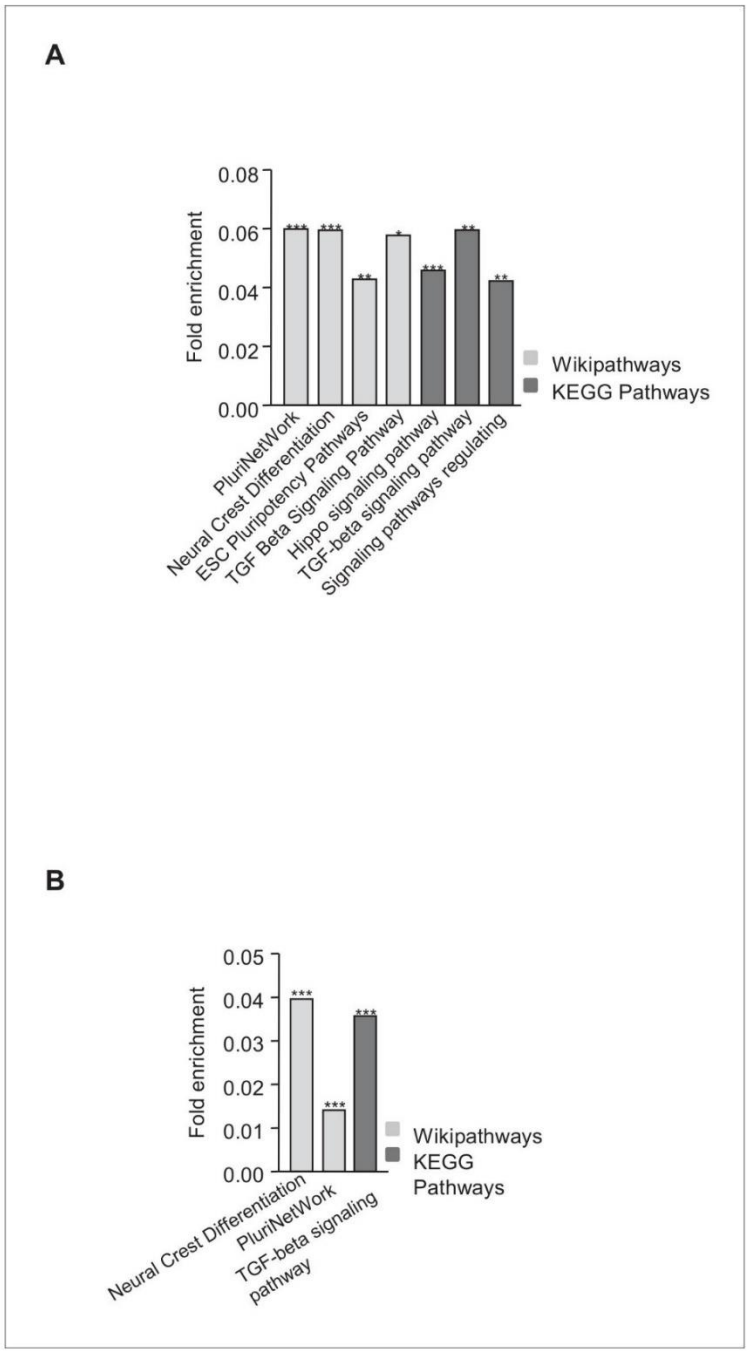


Figure 3—figure supplement 1. Functional analysis of genes downregulated in the absence of TRF1 that show recruitment of SUZ12 and H3K27me3. (**A**) KEGG and Wikipathways analyses of genes downregulated in the absence of TRF1 that show recruitment of SUZ12 showed that these genes were significantly enriched in important pathways controlling pluripotency and differentiation. (**B**) KEGG and Wikipathways analysis of genes downregulated in the absence of TRF1 that show recruitment of H3K27me3 showed that these genes were

Figure 3—figure supplement 1 continued on next page

Figure 3—figure supplement 1 continued

significantly enriched in important pathways controlling pluripotency and differentiation. * = p value<0.05; ** = p value<0.01; *** = p value<0.001.

DOI: https://doi.org/10.7554/eLife.44656.012

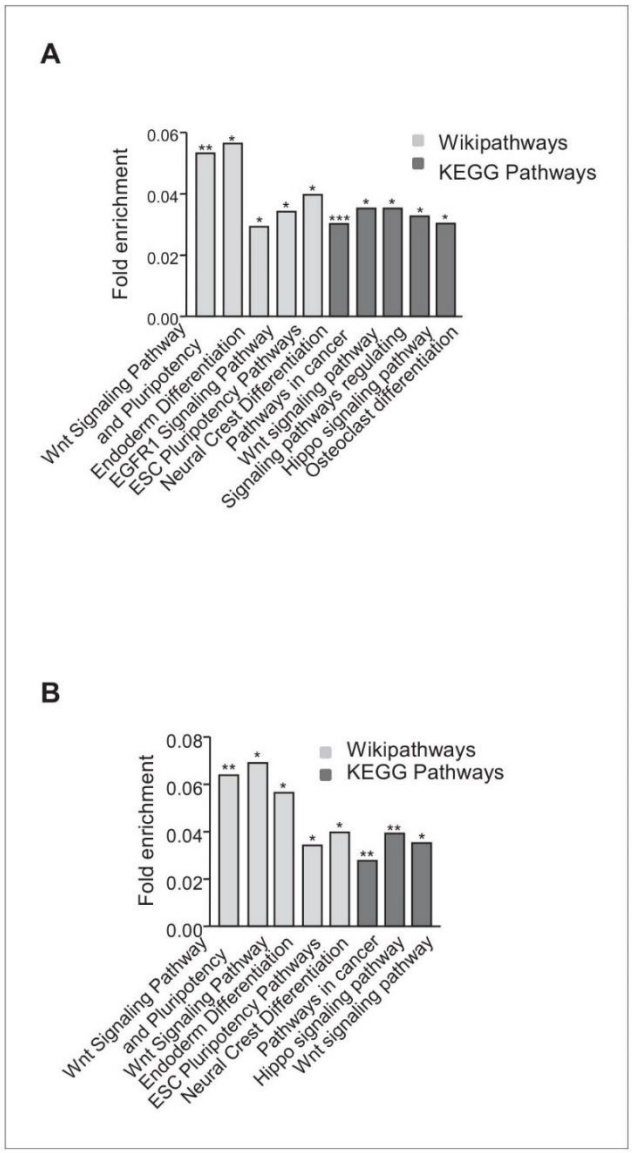


Figure 4—figure supplement 1. Functional analysis of genes that are upregulated in the absence of TRF1 and that show recruitment of SUZ12 and H3K27me3. (**A**) KEGG and Wikipathways analysis of genes that are upregulated in the absence of TRF1 that show recruitment of SUZ12. The histogram shows that these genes are significantly enriched in important pathways controlling pluripotency and differentiation. (**B**) KEGG and Wikipathways analysis of genes that are upregulated in the absence of TRF1 and that show recruitment of H3K27me3 . The histogram shows that these genes are significantly enriched in important pathways controlling pluripotency and differentiation.

DOI: https://doi.org/10.7554/eLife.44656.014

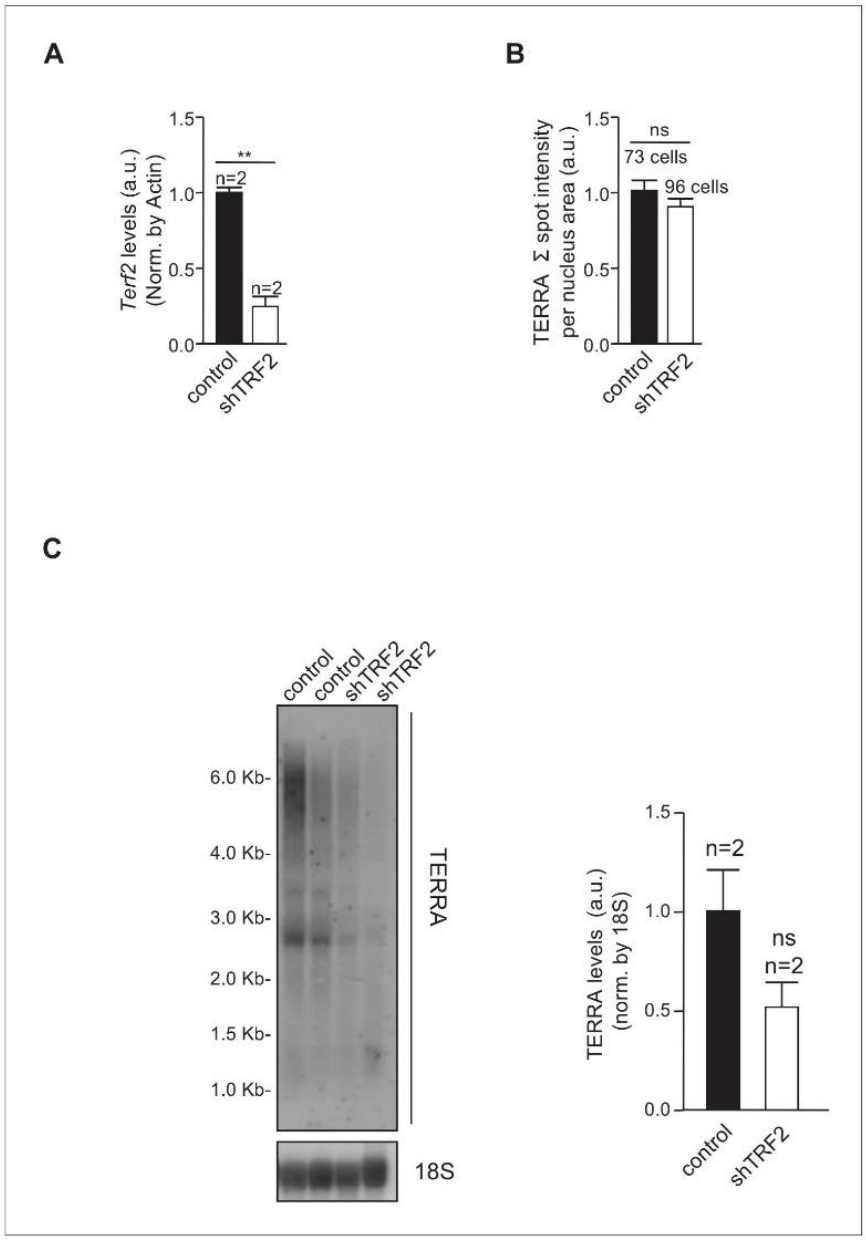


Figure 5—figure supplement 1. TRF2 abrogation in *Trp53*$^{-/-}$ 2i-grown iPS cells does not change TERRA levels. (**A**) *Trf2* expression levels, as measured by q-PCR, in iPS cells infected with a scrambled (control) shRNA or an shRNA against TRF2. Note the clear reduction in *Trf2* levels. N = number of independent experiments.; Error bars = SE. Statistical analysis was carried out by Student's t-test. ** = Adjusted p-value <0.001. (**B**) TERRA levels in control cells or cells in which TRF2 is abrogated, as measured by RNA-FISH. Note that TERRA levels do not change when TRF2 expression is reduced. N = number of cells analyzed. Error bars = SE. Statistical analysis was carried out by Student's t-test. (**C**) TERRA levels in control cells or in cells in which TRF2 is abrogated, as measured by Northern Blot. Note again that the depletion of TRF2 does not increase levels of TERRA RNA. N = number of independent experiments. Error bars = SE. Statistical analysis was carried out by Student's t-test.
DOI: https://doi.org/10.7554/eLife.44656.016

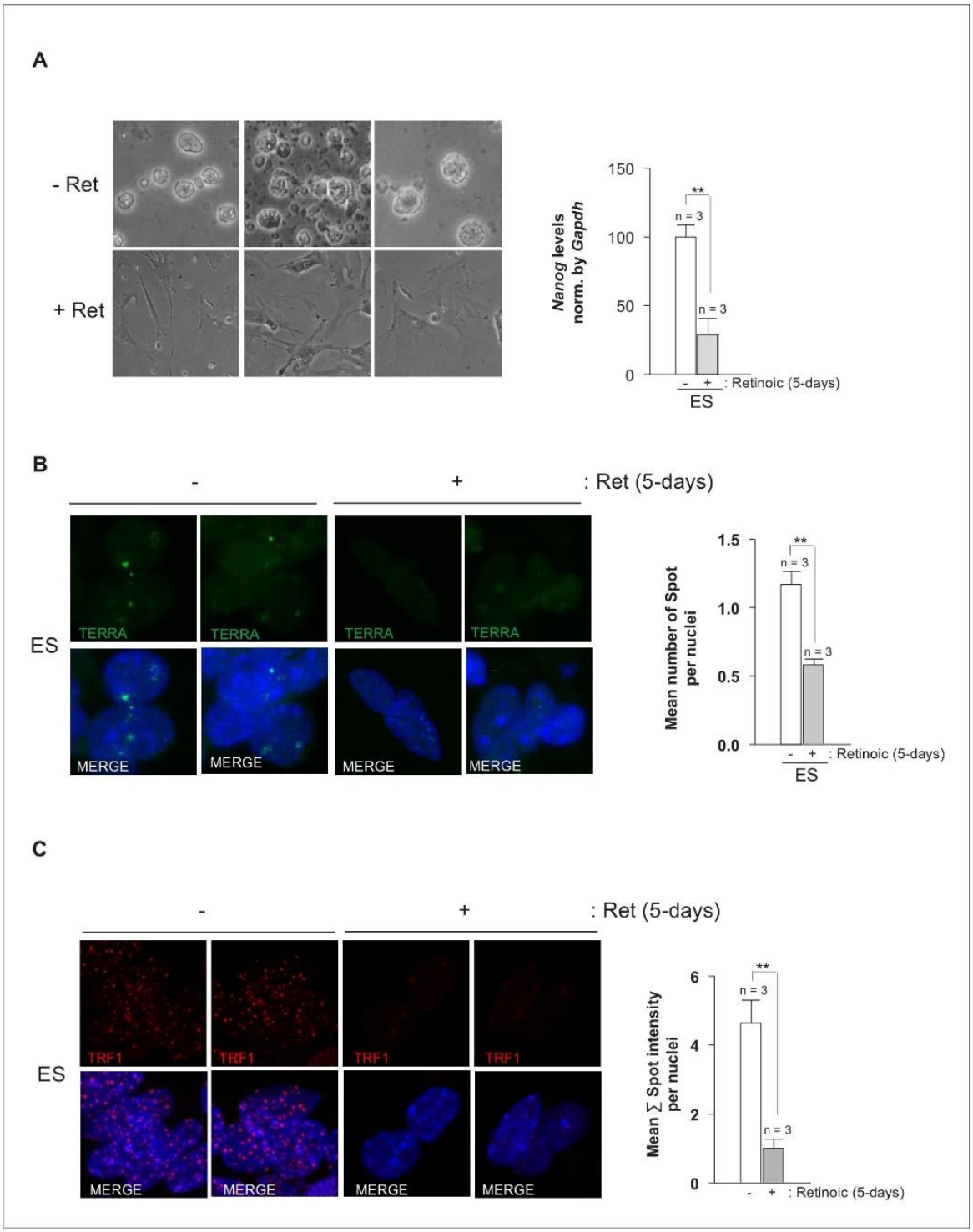


Figure 5—figure supplement 2. TRF1 protein and TERRA levels decrease upon differentiation of ES cells. (**A**) Left, images of ES cells that are untreated (–Ret) or treated (+Ret) with retinoic acid to induce differentiation. Note that cells with retinoic acid show a clear differentiated morphology

Figure 5—figure supplement 2 continued on next page

Figure 5—figure supplement 2 continued

when compared to untreated cells. Right, quantification of *Nanog* levels as measured by q-PCR. Note the clear decrease of *Nanog* expression, confirming the differentiation of the cells. (**B**) Levels of TERRA in untreated cells and in cells treated with retinoic acid to induce differentiation, as measured by RNA-FISH. Left, representative images. Right, quantification of TERRA levels. Note the significant decrease of TERRA in differentiated cells. N = number of independent experiments. Error bars = SE. Statistical analysis was carried out by Student's t-test. ** = Adjusted p-value<0.001. (**C**) Left, representative images of immunofluorescence against TRF1 protein (red) of untreated ES cells (-) or ES cells that have been treated with retinoic acid (+) for 5 days. Right, quantification of TRF1 levels per nuclei. n = number of independent experiments. Error bars = SE. Statistical analysis was carried out by Student's t-test. ** = Adjusted p-value <0.001.

DOI: https://doi.org/10.7554/eLife.44656.017

Anexo 2

Otros artículos realizados durante la tesis doctoral

Anexo 2: Safety of whole-body abrogation of the TRF1 shelterine protein in wild-type and cancer-prone mouse models